공학석사학위논문

Evaluation of Flood Risk Using Bayesian Networks

베이지안 네트워크를 이용한 홍수위험도 평가

2012년 8월

서울대학교 대학원

건설환경공학부

선우 우 연

Abstract

Evaluation of Flood Risk Using Bayesian Networks

Woo Yeon Sunwoo

Department of Civil and Environmental Engineering

The Graduate School

Seoul National University

Extreme weather caused by climate change is one of the biggest concerns in the world: natural disasters such as even cold wave and heavy snow, heavy rainfall and flood, drought and forest fire occurred in the world. So as to Korea, the domestic cases of extreme weather are the frequent local rainfall and flooding of downtown Seoul. Many areas are exposed to the increasing flood risk due to urbanization and regional development. Accordingly, people's interest of the relevant water control information has significantly increased and the public's participation in water management and awareness about the environment grew. As a result, it is a trend that the validity of the policy should be supported by convincibility. Therefore, it is important to establish countermeasure to cope with the natural disasters by extreme weather. The various information systems are required to provide

planning and prioritizing related to water control projects.

Floods are a typical natural disaster in Korea, so flood risk analysis was carried out. We explored the impact of complex causal relationship with climate characteristics as well as socio • economic characteristics. Bayesian networks were applied to reflect flood adaptation capacity for our society. Bayesian networks express complex issues including several dependencies and uncertainties, and provide a framework that can model and analyze dependencies of proxy variables associated with the flood risk. In addition, Bayesian networks can take full advantage of a priori knowledge such as expert opinion or literature so that it is effective for analysis of the flood risk that has difficulties in obtaining objective information.

Since the flood damage is estimated by loss of lives or property damage in the present, flood risk was obtained based on the cost of property damage by flood. For evaluation of obtained flood risk a network was constructed for the whole country consisting of 12 cities and provinces. Particularly, a separate network was constructed for Seoul considering its distinctive humanistic and social characteristics from other regions. In addition, to eliminate the uncertainties of input data of proxy variables selected for the flood risk analysis, raw data were used.

The proposed approach mainly consists of three parts: flood risk modeling, flood risk analysis and flood risk diagnosis. First, in the flood risk modeling stage, Bayesian networks are constructed to express the dependencies of proxy variables affecting the flood risk, then the degree of dependencies are represented quantitatively by using conditional probability. Second, the two countermeasures are proposed based on the constructed Bayesian networks. On the one hand, current state of the flood risk is analyzed by calculating the occurrence probability of flood risk. To evaluate the validity of Bayesian networks result is compared with actual flood damage. On the other hand, through a flood risk analysis the influence of a specific proxy variable is considered and results that can quantify the influence are presented. Finally, the national flood risk mapping is presented based on the results of flood risk analysis to diagnose flood risk.

In Seoul, proxy variables of road area ratio, total population, summer precipitation had sensitivity to flood risk. This reflected that the characteristic of urbanization and large population in Seoul. In Korea, number of civil servants related to water, area ratio with the banks, summer precipitation, and capacity of drainage facilities showed significant sensitivity to flood risk. The higher values of above proxy variables, the higher flood risk. Dajeon, Chungcheong-do and Ulsan showed higher flood risk than other area. And especially flood risk of big cities such as Incheon, Busan, Daegu, Gwangju and Daejeon showed higher accuracy with actual flood damage.

The proposed method is expected to support flood risks management effectively by modeling flood risk and diagnosis comprehensively. Especially

regional flood risk results considering the uncertainty analysis are expected to be useful as a basis for establishing the optimal flood protection countermeasures.

keywords: Bayesian networks, flood risk analysis, flood risk mapping

***Student Number*: 2010-23308**

Table of Contents

List of Tables

List of Figures

List of Symbols

Latin Uppercase

A	event
B	event
C	set of children of X_i
E	some set of evidence variables (=$\{E_i^-, E_i^+\}$)
E_i^-	part consisting of assignments to variables in the subtree rooted at X_i
E_i^+	rest of part consisting of assignments to variables in the subtree rooted at X_i
E_j	evidence of X_j
H	entropy function
$H(T\|Y)$	average residual uncertainty
$I(T,Y)$	Shannon's mutual information; sensitivity
P	local probability distribution, $\Pr(X_i = x_i \| Pa_i = pa_i)$
$\mathbf{Pa}_i$	parents of node X_i as well as random variables corresponding to those parents

S	network structure
T	target random variable (flood risk)
X_c	child node of X_i
X_i	random variable corresponding to i 's node ($i = 1, 2, ..., n$)
$\mathbf{X}$	set of $\{X_1, ..., X_n\}$
X_j	case of several children of X_i
X_p	parent node of X_i
Y	random variable: proxy variable
Z	random variable

Latin Lowercase

i	index ($= 1, 2, ..., n$)
j	number of cases of X_c
n	number of variables
$\mathbf{pa}_i$	values of random variables $\mathbf{Pa}_i$ ($i = 1, 2, ..., n$)
r	correlation coefficient
t	value of target random variable T
x_i	value of random variable X_i ($i = 1, 2, ..., n$)

$\bar{x}$	mean value of X
$\mathbf{x}$	set of $\{x_1, ..., x_n\}$
y	value of random variable Y
$\bar{y}$	mean value of Y
z	value of random variable Z

Greek Uppercase

$\mathbf{\Pi}_i$	subset of $\{X_1, ..., X_{i-1}\}$

Greek Lowercase

α	constant independent of X_i
θ	state of nature, parameter
$\boldsymbol{\theta}$	vector of unknown parameter
ϕ	null set
$\lambda(E_i)$	$\Pr(E_i^- \mid X_i)$: conditional probability of E_i^- given X_i
$\lambda_j(X_i)$	$\Pr(E_j^- \mid X_j)$: conditional probability of E_j^- given X_j
$\boldsymbol{\pi}_i$	values of $\mathbf{\Pi}_i$

$\pi(X_i)$	$\Pr(X_i \mid E_i^+)$: conditional probability of X_i given E_i^+
$\pi_j(X_p)$	$\Pr(X_p \mid E)$: conditional probability of X_p given E over $\lambda_j(X_p)$

Special Symbols

$\mathbf{X} \setminus \mathbf{Y}$	values in $\mathbf{X}$ that are not in $\mathbf{Y}$
Pr	probability that some event will occur
$\prod$	product
$\sum$	summation
$\cap$	intersection

1. Introduction

1.1 Research background

As the features of recent flood are spatially concentrated and show heavy rainfalls in a very short time, loss of life and property damage increase greatly by the impact of climate change and unpredictable abnormal changes of weather. The extent of damage by flood increases as use of riparian land is increased by development of economy and society. The extent of risk increases in many regions exposed to flood by heavy rainfalls occurred frequently because of urbanization and community development. Therefore public interest in water control information is dramatically increased and socially reasonable justification of water control policy is needed by increment of public participation to the water control and further raising awareness about environment.

For clear and reliable decision making of water control policies, international organizations like UN (United Nations), World Bank and IADB (Inter-American Development Bank), and many research centers have developed water control index that can be used easily by general persons and decision makers in water resources related fields. In Korea, the nation's water resource management process provides various information of water control by WAMIS (Water Management Information System), however provided categories limited to only basic data such as characters and figures so that political decision makers, water resource experts and general person cannot obtain proper information for simultaneous analysis of status.

1.2 Research objectives and methodology

1.2.1 Objectives

The variables used for flood risk analysis are a mean for measuring change of the flood control status, and providing the priority to resolve the flood damage problem. It has advantages such as inspecting the present status of administrative region and basin to deal with a flood control project and, providing a reasonable decision making for flood control policies. It is needed to evaluate the flood risk using intuitive variable to let people know the status of flood control easily for public response in establishing flood control policy. In addition, it is necessary to construct a system for providing flood control information to users who work for the flood control field in various forms. Recently, there is also the effort of government for information-oriented system for water control. Methods were suggested to connect water information, effective management and maintenance of information system in research of basic strategy setting for water resources informatization (Ministry of Construction and Transportation, 2006). Relating to information-oriented system for water control, not only basic information about water resource but also various forms of information by analyzing and post processing the basic information are needed.

As needs of the quantitative analysis research rises in addition to methods using conventional index and indicator of flood characteristic, research of flood risk is actively in process. Basic structures of the index are theme frame work and cause-effect chain framework. Theme framework assorts a phenomenon by fields then departmentalizes into sub-fields by political purpose in that field and selects related index. Cause-effect chain framework is a framework that constructs index by analyzing inter-relationships among a phenomenon to be measured, causes that made the result and efforts to make the phenomenon stronger or weaker. Research for cause-effect chain framework about developed indexes for water management is inadequate in Korea. That is, most researches use unconditional approach methods. Moreover in most researches flood risk is just evaluated by standardization of selected variables.

Although indexing process digitizing the related variables has merit to make judgment of flood control status totally, it also has limit of difficulty for comparing of the specific items in detail because value of the variable cannot reflect the absolute variable. Unless applying of variables related to flood risk also is oriented from causal relationships it will be difficult to understand inter-relation between causal factor and damage of flood. Flood risk analysis should be applied to interdependence of the selected variables.

In this research, to consider complex causal relationships among climate characteristics as well as socio • economic characteristics, Bayesian networks (BNs) were applied to reflect flood adaptation capacity for our society. Bayesian networks express complex issues including several dependencies and uncertainties, and provide a framework that can model and analyze dependencies of proxy variables associated with the flood risk.

1.2.2 Methodology

For flood risk analysis a network was constructed for the country consist of 12 cities and provinces. Especially, a separate network was configured for Seoul considering its differentiating humanistic and social characteristics from other regions. In addition, to eliminate the uncertainties of input data of proxy variables, raw data were used. To consider casual relationships and interdependencies among proxy variables used for flood risk analysis Bayesian networks were applied. The procedure developed for flood risk mapping used mainly three steps: flood risk modeling, flood risk analysis and flood risk diagnosis.

Step 1: Flood risk modeling

Step 2: Flood risk analysis

Step 3: Flood risk diagnosis

Step 1 consisted of 3 sub-steps which were 1) selection of proxy variables, construction of the database by actual data and 2) separation of proxy variables into two types as discrete or continuous and then dividing to three (low, normal and high) ranges for continuous type. 3) Last sub-step was construction of flood risk model using Netica^{TM} 4.16. (Norsys Software Corporation: www.norsys.com). Netica^{TM} is Netica is a powerful, easy-to-use, complete program for working with belief networks and influence diagrams. It has an intuitive and smooth user interface for drawing the networks, and the relationships between variables may be entered as individual probabilities, in the form of equations, or learned from data files.

Bayesian networks were constructed to express the dependencies of nodes (proxy variables) affecting the flood risk as an arcs, then the degree of dependencies quantitatively represented by using Conditional Probability Tables (CPTs). The constructed networks were modified iteratively by the Baye's theorem and data analysis. Above 3 sub-steps of step 1 were needed for structure identification process.

In Step 2, the two countermeasures were proposed based on Bayesian networks (BNs) constructed. On the one hand, the flood risk by calculating the occurrence probability of flood risk was compared with actual flood damage in order to evaluate the validity of the BNs. On the other hand, through a flood risk analysis the influence of a specific proxy variable was considered and results that could quantify the influence were presented.

Finally in Step 3, the flood risk mapping was developed based on the results of flood risk CCGISTM (Climate Change adaptation toolkit based on GIS, National Institute of Environmental Research) to classify flood risk. CCGIS^{TM} is a tool that can calculate vulnerability index of climate change to establish adaptation measures and impact assessment of climate change. CCGIS^{TM} has been built based on GIS information. The procedure for application in this study can be shown in Figure 1.1.

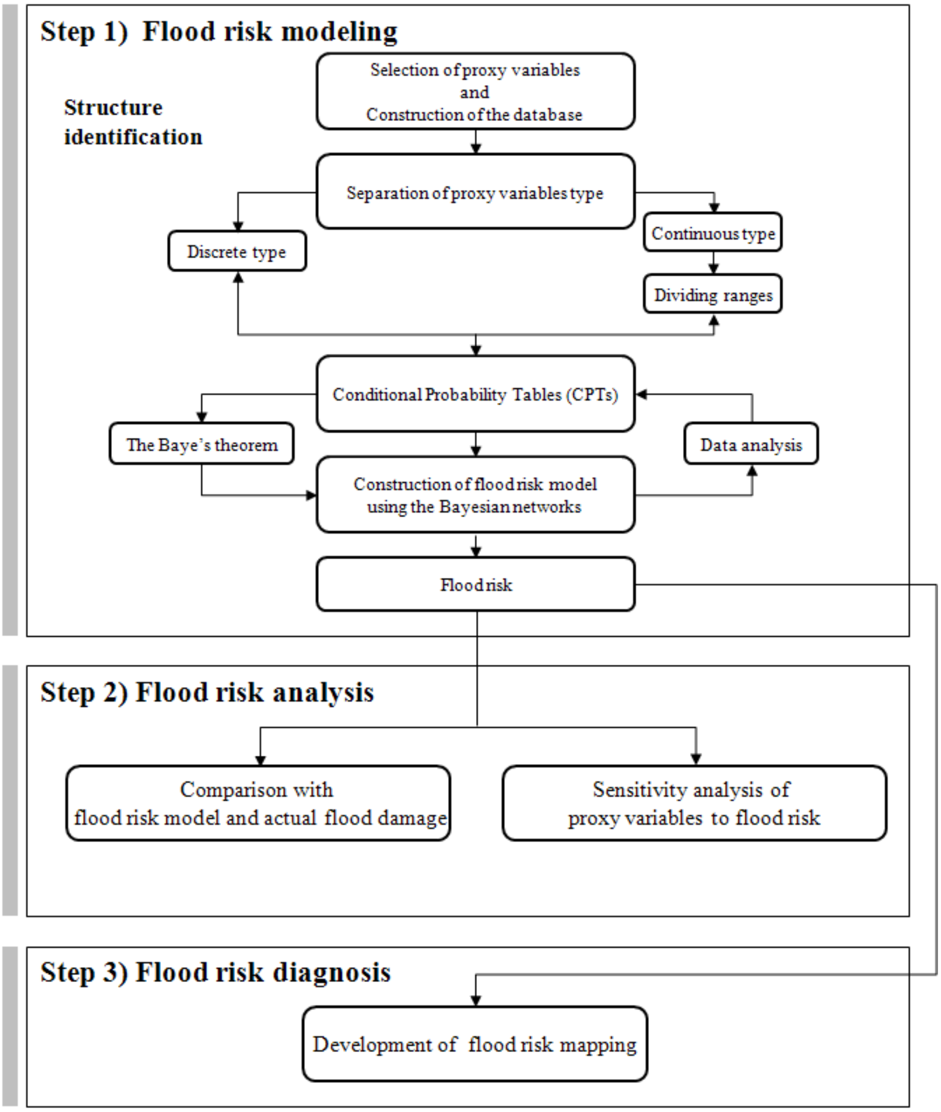


Figure 1.1 Procedure of the research

1.3 Previous research

1.3.1 Hazard mapping and flood risk analysis

Today characteristics of flood risk represent higher and higher intensity due to climate change and unpredictable abnormal changes of weather. To integrate local knowledge, GIS and maps are generally used for the process of disaster risk management. Hazard mapping is one of the first steps of producing a community risk assessment (Noson, 2002). Zhang et al. (2002) calculated the flood damage risk by adding factors such as meteorological triggering factor, natural and socio-economic factors contributing to flood damage generating, and historical flood damages etc. Maps can provide clear, attractive pictures of the geographic distribution of potential risks that can be appreciated by local people with no specialist knowledge. Hazard maps are fundamental to the development of a community-based methodology for collecting and displaying the disaster vulnerabilities and risks that comprise the core content of local knowledge (Hatfield, 2006). Han et al. (2007) suggested mapping method of flood risk from flood hazard mapping by flood risk index and flood risk intensity applying risk concept quantitatively.

In Korea, research for flood risk analysis is in progress by index estimation related to flood and flood risk mapping. Lee et al. (2006) made evaluation model of regional safety for Seoul by classifying flood damage index into primary factors of hazard and adaptation. Kim et al. (2007) set flood risk index by calculating direct runoff of water resources unit region for Yongsan River. Jang and Kim (2009) suggested a methodology to estimate the potential property loss and assessed the flood risk using a regional regression analysis. Lim et al. (2010) applied flood risk index to whole country by classification to pressure index, state index and response index. National Institute of Environmental Research (2012) calculated climate change vulnerability index of 232 regional municipal corporations considering water control and water use using 21 proxy variables. In statistics, a proxy variable is something that is probably not in itself of any great interest, but from which a variable of interest can be obtained. In order for this to be the case, the proxy variable must have a close correlation, not necessarily linear or positive, with the inferred value. The importance of each proxy variable is expressed weighted value. Likewise, the flood risk research is being actively discussed not only Korea but also other foreign countries. Karmakar and Simonovic (2007) defined flood risk as multiplication of hazard, vulnerability and exposure. They also defined exposure index as classification of physical vulnerability, economical vulnerability and vulnerability for infrastructure. Kannami (2008) defined flood hazard index as multiplication of hazard exposure and basic vulnerability divided by capacity represented as average of unstructured and structured measures.

However, the flood risk index developed is not enough to decide a specific action plan about flood control project. There are exactly casual relationships among variables used for evaluating flood risk. Therefore methods are required that would handle uncertainty and enable various domain of evidence to be connected. In addition, such methods must be able to integrate multiple system processes.

1.3.2 Bayesian networks (BNs) for risk modeling

Duckstein et al. (1987) outlined how Bayesian methodology could be used to forecast resulting changes in annual flood frequencies. Bayesian networks (Pearl, 1988) have gained a reputation of being powerful techniques for modeling complex problems involving uncertain knowledge. The causal information encoded in Bayesian networks facilitates the analysis of actions, sequences of events, observations, consequences and expected utility. Bernier (1994a) reviewed Bayesian tests for detecting the date of climate change in hydrological time series, and illustrated his methods with applications to the Harricana, St. Lawrence and Senegal rivers and precipitation in the North American Great Lakes region. In Bernier's study (1994b) of droughts in the Senegal River basin, uncertainties concerning the probability distribution of annual flows were analyzed with a Bayesian approach. Since Bayesian networks provide a robust and mathematically coherent framework for modeling uncertain, non-linear and complex domains (Pearl 1988; Daniel et al. 2007), their application in environmental sciences is increasingly gaining popularity (McCann et al. 2006; Pollino et al. 2007; Uusitalo 2007).

In a recent study, Dlamini (2011) applied Bayesian networks for fire risk mapping. Forest fires or wildfires in common with flood events are complex events that take place as a result of both natural processes and anthropogenic factors. For analysis of variables influencing the fire risk, a total of thirteen biophysical and socio-economic explanatory variables were analyzed and processed using a Bayesian networks and GIS to generate the fire risk maps. This is the first attempt to estimate and map fire risk in Swaziland, uses a Bayesian networks to spatially compute fire probabilities as measures of fire risk using explanatory factors derived from remotely sensed data and GIS analysis. Note that there are only few reported applications of Bayesian networks in the field of natural hazard risk assessment.

2. Theoretical Background

2.1 Flood risk

The term ‘flood risk’ describes the probability that a flood might occur or an area might be flooded as affected by incidence of causative factors such as impact of climate. Since the flood damage is estimated by loss of lives or property damage in the present, flood risk was obtained based on the cost of property damage by flood. Therefore through the evaluation of flood risk, we can notice the regional probability of flood risk. It may anticipate potential future events or may focus on historic event.

The Intergovernmental Panel on Climate Change (IPCC) represents the disaster risk from climate change as weather and climate events that can contribute to disaster, as well as the exposure and vulnerability of human society and natural ecosystems. Disaster risk also considers the role of development in trends in exposure and vulnerability, implication for disaster risk, and interactions between disaster and development (IPCC 2012). In this perspective, it is important to consider the risk as well as its own risk and the cause of risk and exposure to the risk of social system, with the ability to adapt to the risk in order to reduce the disaster risk.

2.2 Bayesian statistics

To understand Bayesian networks and associated learning techniques, it is important to understand Bayesian statistics. Scientific experimental or observational results generally consist of sets of data of the general form $\mathbf{X} = \{X_1, ..., X_n\}$, where X_i's are observations of interest. Statistical methods are then typically used to derive conclusions on both the nature of the process which has produced those observations, and on the expected behavior at future instances of the same process. A central element of any statistical analysis is the specification of probability model which is assumed to describe the mechanism which has generated the observed data $\mathbf{X}$ as a function of a parameter $\theta \in \boldsymbol{\theta}$, sometimes referred to as the state of nature. All derived statistical conclusions are obviously conditional on the assumed probability model. Nature is s frequent contributor to the uncertainty, and it has become customary to refer to its aspects of uncertainty as the states of nature (Kottegoda and Rosso, 1997)

Statistic is the study of uncertainty. The field of statistics is based on two major paradigms; conventional and Bayesian. Bayesian methods provide a complete paradigm for both statistical inference and decision making under uncertainty. Also they can be derived from an axiomatic system and provide a coherent methodology which makes it possible to incorporate relevant initial information, and which alleviates many difficulties faced by conventional statistical methods. Bayesian probability contrasts with the frequency probability, in which probability is derived from observed frequencies in defined distributions or proportions in populations. Differences in the two interpretations imply different methods in statistics. For example, the mass of Saturn can be estimated by Bayesian methods. However, probability theory using frequency probability cannot be applied to this problem since the mass of Saturn has a determinate, but unknown value. Its value cannot be represented as a random value from a distribution or population. Thus, while population parameters are treated as unknown constants in conventional statistics, unknown parameters or states of nature are considered as random variable in Bayesian statistics. Similarly, when comparing two hypotheses using the same information, the frequency probability theory would state the rejection or non-rejection of the original hypothesis with a particular degree of confidence, while Bayesian methods would state that one hypothesis was more probable than another or that expected loss associated with one hypothesis was less than the expected loss of another (Kim, 2007). Although there is a history of antagonism between the Bayesian and the frequency, in many applications Bayesian methods are more general and appear to give better results than frequency probability.

Conditional probability, Baye's theorem and chain rule of probability

The probability of one event given that another event occurs is a conditional probability. A conditional probability is defined in terms of the probabilities of the given events and combinations of them. If the probability of an event B depends on the occurrence of an event A, then the probability can be written by $\Pr(B|A)$, read as the probability of B given A or the conditional probability of B given A has occurred. Thus the $\Pr(B|A)$ is given by

$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)} \tag{2.1}$$

If event B is independent of A, then $\Pr(B|A) = \Pr(B)$. Therefore, the joint probability of two independent events is the product of their individual probabilities.

$$\Pr(A \cap B) = \Pr(A)\Pr(B) \tag{2.2}$$

Here, A and B are disjoint or mutually exclusive if $A \cap B = \phi$. Its occurrence is accompanied by the occurrence of other events $B_1, B_2, ...$ such that the probability of A will depends on which of the event B_i has occurred. In such case, the probability of A will be an expected probability, that is, the average probability weighted by those of B_i. If event B_i are a set of mutually exclusive and collectively exhaustive, one can determine the probability of another event A from

$$\Pr(A) = \sum_{i=1}^{n} \Pr(A|B_i)\Pr(B_i) \tag{2.3}$$

This is called as the theorem of total probability, which can be derived by the definition of conditional probability. Bayes' theorem which is called after Thomas Bayes, an English cleric and philosopher of the eighteenth century provides a definite relationship between the two events. This theorem is a result in probability theory, which relates the conditional and marginal distributions of random variables. By rewriting Equation (2.1) and substituting from Equation (2.3) for $\Pr(A)$, then Bayes' theorem is represented by

$$\Pr(B_i | A) = \frac{\Pr(B_i)\Pr(A | B_i)}{\sum_{i=1}^{n} \Pr(A | B_i)\Pr(B_i)} \tag{2.4}$$

The chain rule of probability permits the calculation of any member of the joint distribution of a set of random variables using only conditional probabilities. Consider an indexed set of $X_1, ..., X_n$. To find the value of this member of the joint distribution, we can apply the definition of conditional probability to obtain

$$\Pr(X_1, ..., X_n) = \Pr(X_n | X_{n-1}, ..., X_1) \cdot \Pr(X_{n-1}, ..., X_1) \tag{2.5}$$

Repeating this process with each final term creates the product

$$\prod_{i=1}^{n} \Pr(X_i | X_1, ..., X_{i-1}) \tag{2.6}$$

The chain rule is useful in the study of Bayesian networks, which describe a probability distribution in terms of conditional probabilities.

2.3 Bayesian networks

Bayesian networks (BNs) are a graphical model for probabilistic relationships among a set of variables. Over the last decade, Bayesian networks have become a popular representation for encoding uncertain expert knowledge in expert systems (Heckerman, 1995). More recently, researchers have developed methods for learning Bayesian networks from data. The techniques that have been developed are new and still evolving, but they have been shown to be remarkably effective for some data-analysis problems. The network is graphically represented as a Directed Acyclic Graph (DAG) consisting of nodes and arcs where the nodes stand for random variables and the arcs show dependency between random variables. The causal information encoded in Bayesian networks facilitates the analysis of actions, sequences of events, observations, consequences, and expected utility (Pearl 1988). A BN can be described as a graphical model that efficiently encodes the joint probability distribution (physical or Bayesian) for a large set of variables.

In order for a Bayesian networks to model a probability distribution, the following must be true by definition: Each variable is conditionally independent of all its non-descendants in the graph given the value of all its parents. In simple case, suppose random variable Z is parent of Y, and Y

is parent of X . X and Z are independent, given Y .

$$\Pr(X \mid Y, Z) = \Pr(X \mid Y) \tag{2.7}$$

We know that the value of all of X 's parent (namely, Y), and Z is not a descendant of X , X is conditionally independent of Z . Also since independent is symmetric,

$$\begin{aligned}
\Pr(Z \mid Y, X) &= \frac{\Pr(X, Y \mid Z)\Pr(Z)}{\Pr(X, Y)} && \text{(Bayes' theorem)} \\
&= \frac{\Pr(Y \mid Z)\Pr(X \mid Y,\ Z)\Pr(Z)}{\Pr(X \mid Y)\Pr(Y)} && \text{(Chain rule)} \\
&= \frac{\Pr(Y \mid Z)\Pr(X \mid Y)\Pr(Z)}{\Pr(X \mid Y)\Pr(Y)} && \text{(By Equation 2.7)} \\
&= \frac{\Pr(Y \mid Z)\Pr(Z)}{\Pr(Y)} = \Pr(Z \mid Y). && \text{(Bayes's theorem)}
\end{aligned} \tag{2.8}$$

A Bayesian networks for a set of random variables $\mathbf{X} = \{X_1, \ldots, X_n\}$ consists of (1) a network structure S that encodes a set of conditional independence assertions about variables in $\mathbf{X}$, and (2) a set P of local probability distributions associated with each variable. Together, these components define the joint probability distribution for $\mathbf{X}$. The network structure S is a directed acyclic graph. The nodes in S are in one-to-one correspondence with the variables $\mathbf{X}$. We use X_i to denote both the variable and its corresponding node, and $\mathbf{Pa}_i$ to denote the parents of node X_i in S as well as the variables corresponding to those parents. The lack of possible arcs in S encodes conditional independencies. In particular, given structure S, joint probability distribution for $\mathbf{X}$, general product (chain) rule for Bayesian networks is given by

$$\Pr(\mathbf{X} = \mathbf{x}) = \prod_{i=1}^{n} \Pr(X_i = x_i \,|\, \mathbf{Pa}_i = \mathbf{pa}_i) \tag{2.9}$$

The local probability distributions P are the distributions corresponding to the terms in the product of Equation (2.9). Consequently, the pair (S, P) encodes the joint probability distribution $\Pr(\mathbf{X} = \mathbf{x})$.

As process of building a Bayesian networks we must (1) correctly identify the goals of modeling, (2) identify many possible observations that may be relevant to the problem, (3) determine what subset of those observations is worthwhile to model, and (4) organize the observations into variables having mutually exclusive and collectively exhaustive states. Difficulties here are not unique to modeling with Bayesian networks, but rather are common to most approaches. Although there are no clean solutions, some guidance is offered by decision analysts (Howard and Matheson, 1983) and (when data are available) statisticians (Tukey, 1977).

In the next phase of Bayesian networks construction, we build a directed acyclic graph that encodes assertions of conditional independence. One approach for doing so is based on the following observations. From the chain rule of probability, we have

$$\Pr(\mathbf{X} = \mathbf{x}) = \prod_{i=1}^{n} \Pr(X_i = x_i \mid X_1 = x_1, \ldots, X_{i-1} = x_{i-1}) \tag{2.10}$$

Now, for every X_i, there will be some subset $\mathbf{\Pi}_i \subseteq \{X_1, \ldots, X_{i-1}\}$ such that X_i and $\{X_1, \ldots, X_{i-1}\} \setminus \mathbf{\Pi}_i$ are conditionally independent given $\mathbf{\Pi}_i$. That is,

$$\Pr(X_i = x_i \mid X_1 = x_1, \ldots, X_{i-1} = x_{i-1}) = \Pr(X_i = x_i \mid \mathbf{\Pi}_i = \boldsymbol{\pi}_i) \tag{2.11}$$

Combining Equations (2.10) and (2.11), we obtain

$$\Pr(\mathbf{X} = \mathbf{x}) = \prod_{i=1}^{n} \Pr(X_i = x_i \mid \mathbf{\Pi}_i = \boldsymbol{\pi}_i) \tag{2.12}$$

Comparing Equations (2.9) and (2.12), we see that the variables sets $(\mathbf{\Pi}_1,...,\mathbf{\Pi}_n)$ correspond to Bayesian networks parents $(\mathbf{Pa}_1,...,\mathbf{Pa}_n)$, which in turn fully specify the arcs in the network structure *S*. Consequently, to determine the structure of a Bayesian networks we (1) order the variables somehow, and (2) determine the variables sets that satisfy Equation (2.11) for $i = 1,...,n$. This approach has a serious drawback. If we choose the variable order carelessly, the resulting network structure may fail to reveal many conditional independencies among the variables. Thus, in the worst case, we have to explore $n!$ variable orderings to find the best one. Fortunately, there is another technique for constructing Bayesian networks that does not require an ordering.

For decomposing the probabilities, suppose $\Pr(X_i|E)$ where E is some set of evidence variables. Let's split E into two parts: E_i^- is the part consisting of assignments to variables in the subtree rooted at X_i. E_i^+ is the rest of it.

$$
\begin{aligned}
\Pr(X_i \mid E) &= \Pr(X_i \mid E_i^-, E_i^+) \\
&= \frac{\Pr(E_i^- \mid X_i, E_i^+)\Pr(X_i \mid E_i^+)}{\Pr(E_i^- \mid E_i^+)} \\
&= \frac{\Pr(E_i^- \mid X_i)\Pr(X_i \mid E_i^+)}{\Pr(E_i^- \mid E_i^+)} \\
&= \alpha\,\pi(X_i)\,\lambda(X_i)
\end{aligned} \tag{2.13}
$$

where, α is a constant independent of X_i, $\pi(X_i)$ is $\Pr(X_i \mid E_i^+)$, $\lambda(X_i)$ is $\Pr(E_i^- \mid X_i)$. First, calculating $\lambda(X_i)$ for non-leaves case, suppose X_i has one child X_c.

$$
\begin{aligned}
\lambda(X_i) = \Pr(E_i^- \mid X_i) &= \sum_j \Pr(E_i^-, X_c = j \mid X_i) \\
&= \sum_j \Pr(X_c = j \mid X_i)\Pr(E_i^- \mid X_i, X_c = j) \\
&= \sum_j \Pr(X_c = j \mid X_i)\Pr(E_i^- \mid X_c = j) \\
&= \sum_j \Pr(X_c = j \mid X_i)\lambda(X_c = j)
\end{aligned} \tag{2.14}
$$

Second, suppose X_i has a set of children C. The contribution of each subtree to $\lambda(X_i)$ is independent.

$$\lambda(X_i) = \Pr(E_i^- \mid X_i) = \prod_{X_j \in C} \lambda_j(X_i)$$
$$= \prod_{X_j \in C} \left[\sum_{X_j} \Pr(X_j \mid X_i) \lambda_j(X_i) \right] \tag{2.15}$$

where, $\lambda_j(X_i)$ is contribution to $\Pr(E_i^- \mid X_i)$ of the part of the evidence lying in the subtree rooted at one of X_i's children X_j.

Computing $\pi(X_i)$ is

$$\pi(X_i) = \Pr(X_i \mid E_i^+) = \sum_j \Pr(X_i, X_p = j \mid E_i^+)$$
$$= \sum_j \Pr(X_i \mid X_p = j, E_i^+) \Pr(X_p = j \mid E_i^+)$$
$$= \sum_j \Pr(X_i \mid X_p = j) \Pr(X_p = j \mid E_i^+)$$
$$= \sum_j \Pr(X_i \mid X_p = j) \frac{\Pr(X_p = j \mid E_i^+)}{\lambda_i(X_p = j)}$$
$$= \sum_j \Pr(X_i \mid X_p = j) \pi_i(X_p = j) \tag{2.16}$$

where, $\pi_i(X_p)$ is defined as $\dfrac{\Pr(X_p \mid E)}{\lambda_i(X_p)}$.

The approach is based on two observations: (1) people can often readily assert causal relationships among variables, and (2) causal relationships typically correspond to assertions of conditional dependence. In particular, to construct a Bayesian networks for a given set of variables, we simply draw arcs from cause variables to their immediate effects. In almost all cases, doing so results in a network structure that satisfies the definition Equation (2.9). The causal semantics of Bayesian networks are in large part responsible for the success of Bayesian networks as a representation for expert systems (Heckerman, 1995). In the final step of constructing a Bayesian networks, we assess the local probability distribution(s) $\Pr(X_i = x_i \mid \mathbf{Pa}_i = \mathbf{pa}_1)$.

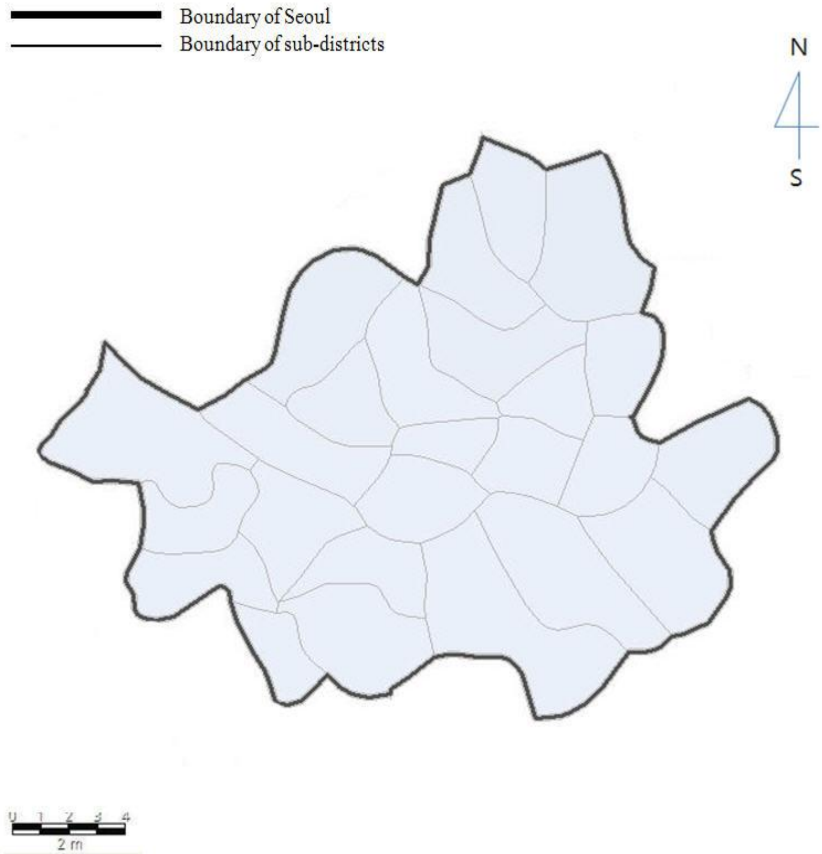
Note that, although we have described these construction steps as a simple sequence, they are often intermingled in practice. For example, judgments of conditional independence and/or cause and effect can influence problem formulation. Also, assessments of probability can lead to changes in the network structure.

3. Network Model Construction

3.1 Study area

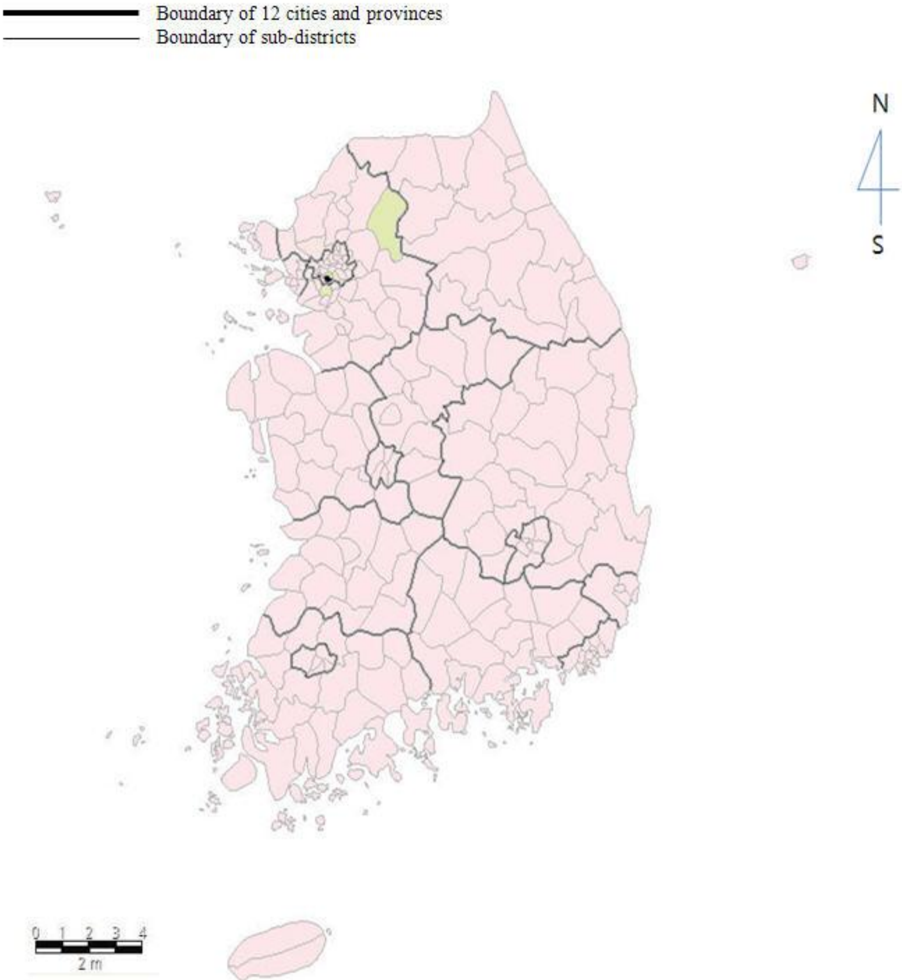
According to the Ministry of Environment in Korea, the frequency and intensity of extreme weather events caused by climate change show short-term change in environment as well as long-term rising temperatures. Korea is located in North-East Asia, has a total land area of 100,033 km^2. The geographic position is between 33° and 43° north latitude and 124° and 131° east longitude (including North Korea). Korea lies in the temperate zone with four distinct seasons. It shows, therefore complex climate characteristics which reveal both continental and oceanic features. It has distinct monsoon wind, a rainy period from the East-Asian Monsoon locally called "Changma", typhoon, and while often heavy snowfalls in winter (Korea Meteorological Administration). Considering the effect of climate change, the study area is selected across the country.

Seoul is the capital and the largest metropolis in Korea. It contains about a quarter of population of Korea. Seoul suffers from a variety of urban environmental problems, including overcrowding, housing shortages,



(a) Seoul

Figure 3.1 Sub-districts map



(b) Korea

Figure 3.1 Sub-districts map (Continued)

Table 3.1 Names and symbols of sub-districts (a) Seoul

No.	Names and symbols of 25 Gu		No.	Names and symbols of 25 Gu	
1	S1	Jongno-Gu	14	S14	Mapo-Gu
2	S2	Jung-Gu	15	S15	Yangcheon-Gu
3	S3	Yongsan-Gu	16	S16	Gangseo-Gu
4	S4	Seongdong-Gu	17	S17	Guro-Gu
5	S5	Gwangjin-Gu	18	S18	Geumcheon-Gu
6	S6	Dongdaemun-Gu	19	S19	Yeongdeungpo-Gu
7	S7	Jungnang-Gu	20	S20	Dongjak-Gu
8	S8	Seongbuk-Gu	21	S21	Gwanak-Gu
9	S9	Gangbuk-Gu	22	S22	Seocho-Gu
10	S10	Dobong-Gu	23	S23	Gangnam-Gu
11	S11	Nowon-Gu	24	S24	Songpa-Gu
12	S12	Eenpyeong-Gu	25	S25	Gangdong-Gu
13	S13	Seodaemun-Gu			

Table 3.1 Names and symbols of sub-districts (b) Korea

No.	Names and symbols of 13 cities and provinces		Number of sub-districts
1	Seoul	S	25
2	Busan	B	16
3	Daegu	DG	4
4	Incheon	I	10
5	Gwangju	GJ	5
6	Daejeon	DJ	5
7	Ulsan	U	5
8	Gyeonggi-do	GG	31
9	Gangwon-do	GW	18
10	Chungcheong-do	C	28
11	Jeolla-do	J	36
12	Gyeongsnag-do	GS	43
13	Jeju	JJ	2

transportation congestion, air pollution, building fires and flooding. For this reason Seoul is an important study area especially for flood risk analysis. The flood risk evaluating units were 25 districts (Gu) for Seoul in Figure 3.2 (a) and 207 cities • counties • districts across the country in Figure 3.2 (b). Names and symbols of sub-districts for Seoul and Korea are in Table 3.1 (a) and (b).

3.2 Proxy variables and data

To evaluate flood risk proxy variables were selected because it is hard to evaluate flood risk directly. Proxy variables consist of mainly three parts as climate exposure, sensitivity to climate exposure and adaptive capacity. Their definitions are shown in Table 3.2. The proxy variables for assessing flood risk are from the research ‘Climate change of department vulnerability mapping’ (National Institute of Environmental Research, 2012) (Table 3.3).

Four variables, daily maximum precipitation, 5 days maximum rainfall period, summer precipitation (June to September), days over 80 mm rainfall would be reflected in flood risk such as concentrated precipitation that causes flood

Table 3.2 Definitions of category of proxy variables

Category	Definition
Climate exposure	The variables representing the impact of climate change (In general climate factors)
Sensitivity to climate exposure	The variables representing the degree of impact of climate exposure (Socio-economic statistical data)
Adaptive capacity	To reduce the impact of climate change variables (Socio-economic statistical data)

Table 3.3 Proxy variables

Category	Variable (node)	Unit	Source (period)
Climate exposure	**Daily maximum precipitation**	mm	National Institute of Environmental Research (1996-2005)
	5 days maximum rainfall period	mm/5 days	
	Summer precipitation (June to September)	mm	
	Days over 80 mm rainfall	Day	
Sensitivity to climate exposure	**Area ratio with the banks**	%	National geographic information institute (2009)
	Total population	persons	Korean statistical information service (2009)
	Regional average slope	degree	Ministry of environment (2009)
	Road area ratio	%	National geographic information institute (2009)
	Population density	persons/km^2	Korean statistical information service (2009)
	Lowland area of less than El. 10 m	km^2	CCGIS™

Table 3.3 Proxy variables (Continued)

Category	Variable (node)	Unit	Source (period)
Adaptive capacity	**Gross Regional Domestic Product (GRDP)**	10^6 won	Korean statistical information service (2009)
	Number of civil servants related to water	persons	Homepage of each local government and telephone response (2011)
	Capacity of drainage facilities	m^3/min	National Watershed Survey Report on five year cycle (2009)

damage (Climate exposure). Another six variables, area ratio with the banks, total population, regional average slope, road area ratio, population density, lowland area of less than El. 10 m could reflect the probability of flood damage, influence factor on flood damage and past flood damage performance (Sensitivity to climate exposure). The other three variables, Gross Regional Domestic Product (GRDP), number of civil servants related to water, capacity of drainage facilities would apply countermeasures for flood control, economic factors and adaptability of community if flooding occurs (adaptive capacity). Daily maximum precipitation, 5 days maximum rainfall period, summer precipitation, days over 80 mm rainfall, area ratio with the banks, regional average slope and GRDP were shown with bold faces to indicate that they were surveyed by other researchers. In order to eliminate the uncertainties of input data of proxy variables selected for the flood risk analysis, raw data were used.

In structure identification of step 1, Equation 3.1 was used to normalize continuous type variables to divide range.

$$\text{Normalization equation} = \frac{\text{Actual value - Minimum value}}{\text{Maximum value - Minimum value}} \times 100 \qquad (3.1)$$

Also, in step 2 the actual flood damage data used to evaluate the validity of the result is re-scaled to compare with the flood risk results from Bayesian network. The re-scaling method is based on ranges of indicators used for comparing of results. The values of actual flood damage data were re-scaled by normalization for the range of 0 ~ 1 by Equation 3.1.

3.3 Flood risk modeling

Bayesian networks allow the prediction of a discrete outcome from a set of variables that may be continuous or discrete type. In order to represent continuous value in Bayesian networks, nodes values are divided into sub-ranges according to existing guidelines and on the basis of expert judgment or using percentiles of data (Marcot et al., 2006; McCann et al., 2006; Uusitalo, 2007). The second step is to determine the relationships among the variables and establish the graphical structure of the model through explicitly displaying the influence diagram or the causal web of interacting variables (Marcot et al., 2006; Uusitalo, 2007). Bayesian networks offer a technique to approach problems with probabilistic conclusions based on probability propagation of evidence through the computation of conditional probability

Tables (CPTs) for each node (random variable) regardless of the number of parents. (Spiegelhalter et al.,1993; Daniel et al,. 2007) Quantitative inference can be achieved based on the parameters of Bayesian networks whereby a joint distribution over all the variables can be computed as a function of the parameters.

$$\Pr(X_1,...,X_n) = \prod_{i=1}^{n} \Pr(X_i \mid \mathbf{Pa}_i) \tag{3.1}$$

where, $\mathbf{Pa}_i$ represents the predictor (parent) variables. Since empirical data are available, the CPTs are estimated or approximated based on the sources of evidence.

Construction of Bayesian networks model was needed to classify the state of proxy variables. According to characteristics of variables types of nodes, proxy variables were classified into discrete or continuous type based on raw data. The discrete type variables are Days over 80 mm rainfall, Lowland area of less than El. 10 m, Number of civil servants related to water and Capacity of drainage facilities. The others were represented as continuous type. The data of discrete type was divided into high (1) or low (0) by the mean value of proxy variable. The range of continuous type nodes values were divided into 3 levels as low, normal and high by criterion values. Basically criterion values used for dividing levels were 1σ but in some nodes values of standard deviation were larger than mean values, so that 1/3 quantile values were used for that nodes instead of 1σ values. Histogram of data for each node were attached to Appendix A. Number of proxy variables used for Seoul was 10 except for GRDP, Number of civil servants related to water and Lowland area of less than El. 10 m because they have same values in Seoul. Input data and their brief descriptions are in Table 3.4 (a) and 3.4 (b).

Table 3.4 Input data and their brief descriptions (a) Seoul

Variable	Node	Type	States (ranges)
Flood	X_1	Discrete	Present (flood), Absent (no flood)
Daily maximum precipitation	X_2	Continuous	< 65.7 mm , 65.7 – 75.2 mm, > 75.2 mm
5 days maximum rainfall period	X_3	Continuous	< 117. 71 mm/5 days, 117. 71 – 139.5 mm/5 days, > 139.5 mm/5 days
Summer precipitation (June to September)	X_4	Continuous	< 426.7 mm, 426.7 – 486.8 mm, > 486.8 mm
Area ratio with the banks	X_5	Continuous	< 0.31 %, 0.31 – 12.4 %, > 12.4 %
Total population	X_6	Continuous	< 271,425 Persons, 271,425 – 509,535 Persons, > 509,535 Persons
Regional average slope	X_7	Continuous	< 2. 6°, 2.6 – 12.9°, > 12.9°
Road area ratio	X_8	Continuous	< 83.6 %, 83.6 – 90.5 %, > 90.5 %
Population density	X_9	Continuous	< 11,995 persons/km^2, 11,995 – 22,305 persons/km^2, > 22,305 persons/km^2

Table 3.4 Input data and their brief descriptions (a) Seoul (Continued)

Variable	Node	Type	States (ranges)
Capacity of drainage facilities	X_{10}	Discrete	Low < mean value < High
Days over 80 mm rainfall	X_{11}	Discrete	Low < mean value < High

Table 3.4 Input data and their brief descriptions (b) Korea

Variable	Node	Type	States (ranges)
Flood	X_1	Discrete	Present (flood), Absent (no flood)
Daily maximum precipitation	X_2	Continuous	< 70 mm, 70 – 85 mm, > 85 mm
5 days maximum rainfall period	X_3	Continuous	< 120 mm/5 days, 120 – 165 mm/5 days, > 165 mm/5 days
Summer precipitation (June to September)	X_4	Continuous	< 490 mm, 490 – 630 mm, > 630 mm
Area ratio with the banks	X_5	Continuous	< 0.9 %, 0.9 – 4.9 %, > 4.9 %
Total population	X_6	Continuous	< 60,000 Persons, 60,000 – 216,000 Persons, > 216,000 Persons
Regional average slope	X_7	Continuous	< 7.2°, 7.2 – 17.1°, > 17.1°
Road area ratio	X_8	Continuous	< 91 %, 91 – 98.7 %, > 98.7 %
Population density	X_9	Continuous	< 250 persons/km^2, 250 – 2,000 persons/km^2, > 2,000 persons/km^2

Table 3.4 Input data and their brief descriptions (b) Korea (Continued)

Variable	Node	Type	States (ranges)
Capacity of drainage facilities	X_{10}	Discrete	Low < mean value < High
Days over 80 mm rainfall	X_{11}	Discrete	Low < mean value < High
Lowland area of less than El. 10 m	X_{12}	Discrete	Low < mean value < High
Number of civil servants related to water	X_{13}	Discrete	Low < mean value < High
GRDP	X_{14}	Discrete	< 25e6 won, 25e6 – 50e6 won, > 50e6 won

Bayesian networks model is to determine the relationships among the variables and establish the graphical structures through explicitly displaying the influence diagram or the causal web of interacting variables. The flood risk network is graphically represented as a directed acyclic graph consisting of nodes and arcs where the nodes stand for proxy variables (random variables) and the arcs show dependency between random variables. The thirteen proxy variables were set up as nodes considering interrelation. The relationships between the proxy variables and flood were derived from a priori knowledge through many previous researches and data analysis of proxy variables. Data analysis was depending on cross correlations between value of proxy variables including actual flood risk (X_{15}) and weighted values of proxy variable (National Institute of Environmental Research, 2012), *etc*. Weighted values of proxy variables are in Table 3.5.

Correlation refers to any of a broad class of statistical relationships involving dependence. The correlation coefficient is useful because it can indicate the degree of connections associated with two variables but is not to explain a causal relationship. A simple but powerful goodness of fit test is the Probability-Plot Correlation Coefficient (PPCC) test developed by Filliben (1975).

Table 3.5 Weighted values of proxy variables

Proxy variable	Node	Weighted value
Daily maximum precipitation	X_2	0.31
5 days maximum rainfall period	X_3	0.19
Summer precipitation (June to September)	X_4	0.11
Area ratio with the banks	X_5	0.07
Total population	X_6	0.10
Regional average slope	X_7	0.11
Road area ratio	X_8	0.07
Population density	X_9	0.12
Capacity of drainage facilities	X_{10}	0.21
Days over 80 mm rainfall	X_{11}	0.23
Lowland area of less than El. 10 m	X_{12}	0.10
Number of civil servants related to water	X_{13}	0.13
GRDP	X_{14}	0.11

The correlation coefficient is calculated as

$$r_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{4.1}$$

The correlation coefficient is +1 in the case of a perfect positive (increasing) linear relationship (correlation), −1 in the case of a perfect decreasing (negative) linear relationship, and some value between −1 and 1 in all other cases, indicating the degree of linear dependence between the variables. As it approaches zero there is less relationship (closer to uncorrelated). The closer the coefficient is to either −1 or 1, the stronger the correlation between the variables. The results of cross correlation between proxy variables including actual flood damage are shown in Table 3.6 (a) and (b) for Seoul and Korea. It was symmetric. Proxy variables and actual flood damage were showed as symbol of nodes from X_2 to X_{15}. The cases of correlation coefficients were 0.5 ~ 1 and -0.5 ~ -1 were shown as shadow zone. And positive correlation coefficients were darker than negative correlation.

Table 3.6 Results of cross correlation coefficients between variables (a) Seoul

Variables	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{15}
X_2	1.000	**0.990**	0.287	-0.253	0.042	0.344	-0.447	0.056	-0.137	0.039	-0.009
X_3	0.990	1.000	0.403	-0.254	0.037	0.349	-0.424	0.062	-0.188	0.126	-0.020
X_4	0.287	0.403	1.000	-0.248	-0.119	0.299	-0.016	-0.080	-0.496	**0.579**	-0.047
X_5	-0.253	-0.254	-0.248	1.000	0.246	**-0.848**	0.411	0.187	0.415	-0.052	0.074
X_6	0.042	0.037	-0.119	0.246	1.000	-0.135	-0.208	0.457	0.239	0.123	**0.582**
X_7	0.344	0.349	0.299	-0.848	-0.135	1.000	-0.577	-0.247	-0.341	0.158	0.091
X_8	-0.447	-0.424	-0.016	0.411	-0.208	-0.577	1.000	0.300	0.010	-0.150	-0.482
X_9	0.056	0.062	-0.080	0.187	0.457	-0.247	0.300	1.000	0.474	-0.178	-0.458
X_{10}	-0.137	-0.188	-0.496	0.415	0.239	-0.341	0.010	0.474	1.000	-0.315	-0.195
X_{11}	0.039	0.126	0.579	-0.052	0.123	0.158	-0.150	-0.178	-0.315	1.000	0.286
X_{15}	-0.009	-0.020	-0.047	0.074	0.582	0.091	-0.482	-0.458	-0.195	0.286	1.000

Table 3.6 Results of cross correlation coefficients between variables (b) Korea

Variables	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
X_2	1.000	**0.911**	**0.608**	-0.021	-0.022	0.096	0.075	0.076	0.083	**0.852**	0.196	0.137	-0.295	-0.059
X_3	0.911	1.000	**0.709**	-0.120	-0.068	0.178	-0.029	-0.019	0.040	**0.825**	0.079	0.066	-0.253	-0.008
X_4	0.608	0.709	1.000	-0.139	-0.217	0.453	-0.240	-0.263	-0.023	**0.621**	-0.241	-0.159	-0.328	0.118
X_5	-0.021	-0.120	-0.139	1.000	0.149	-0.417	0.244	0.107	0.311	-0.068	0.193	0.012	0.020	-0.179
X_6	-0.022	-0.068	-0.217	0.149	1.000	**-0.520**	**0.627**	**0.775**	0.178	-0.134	0.000	0.336	0.206	-0.255
X_7	0.096	0.178	0.453	-0.417	-0.520	1.000	**-0.765**	**-0.604**	-0.216	0.277	-0.303	-0.439	-0.169	0.391
X_8	0.075	-0.029	-0.240	0.244	0.627	-0.765	1.000	**0.902**	0.101	-0.068	0.079	**0.772**	0.076	**-0.700**
X_9	0.076	-0.019	-0.263	0.107	0.775	-0.604	0.902	1.000	0.067	-0.058	-0.056	**0.801**	0.158	**-0.666**
X_{10}	0.083	0.040	-0.023	0.311	0.178	-0.216	0.101	0.067	1.000	0.016	0.041	-0.005	0.186	-0.066

Table 3.6 Results of cross correlation coefficients between variables (b) Korea (Continued)

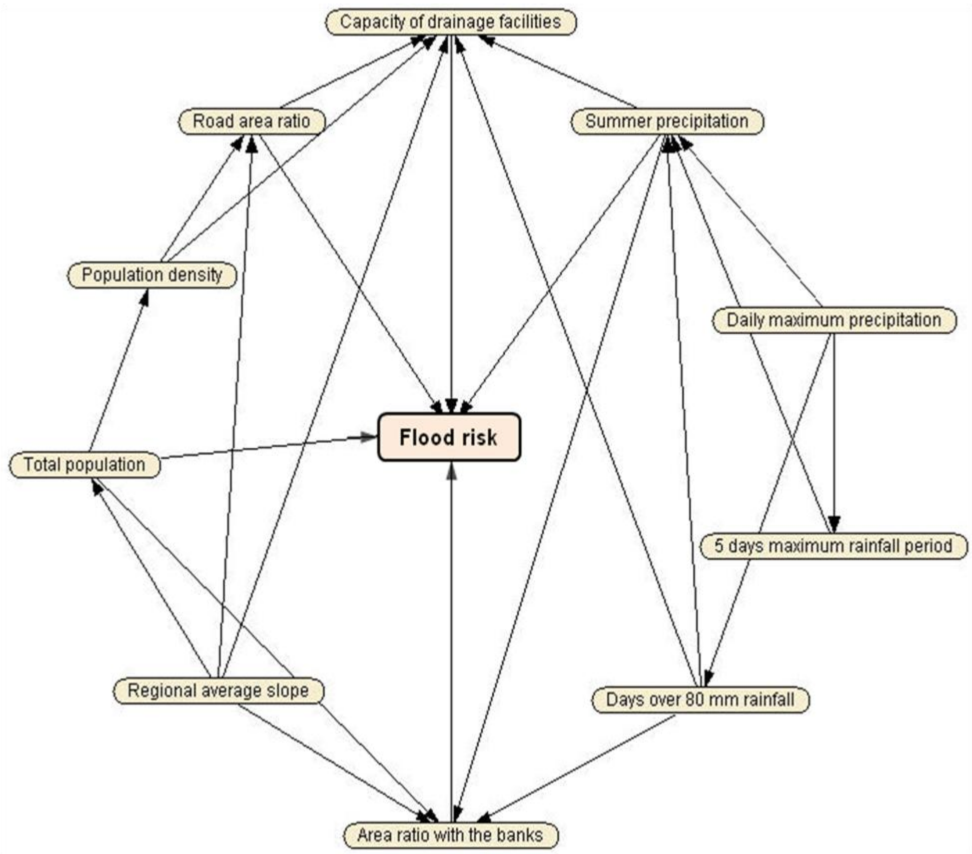
Variables	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
X_{11}	0.852	0.825	0.621	-0.068	-0.134	0.277	-0.068	-0.058	0.016	1.000	0.101	0.034	-0.295	0.030
X_{12}	0.196	0.079	-0.241	0.193	0.000	-0.303	0.079	-0.056	0.041	0.101	1.000	-0.105	-0.015	0.017
X_{13}	0.137	0.066	-0.159	0.012	0.336	-0.439	0.772	0.801	-0.005	0.034	-0.105	1.000	0.049	**-0.723**
X_{14}	-0.295	-0.253	-0.328	0.020	0.206	-0.169	0.076	0.158	0.186	-0.295	-0.015	0.049	1.000	-0.054
X_{15}	-0.059	-0.008	0.118	-0.179	-0.255	0.391	-0.700	-0.666	-0.066	0.030	0.017	-0.723	-0.054	1.000

Using cross correlation liner relationship between variables is analyzed regardless of positive or negative value. In Seoul, total population (X_6), road area ratio (X_8) and population density (X_9) showed high correlations with actual flood damage. The high correlation between daily maximum precipitation (X_2) and 5 days maximum rainfall period (X_3) was 0.990, between area ratio with the banks (X_5) and regional average slope (X_7) was -0.848.

In Korea, number of civil servants related to water (X_{13}), road area ratio (X_8) and regional average slope (X_7) showed high correlations with actual flood damage. Variables with high correlation are reflected as direct arcs in BNs. In Table 3.5, the high weighted values were daily maximum precipitation (X_2), days over 80 mm rainfall (X_{11}) and capacity of drainage facilities (X_{10}).

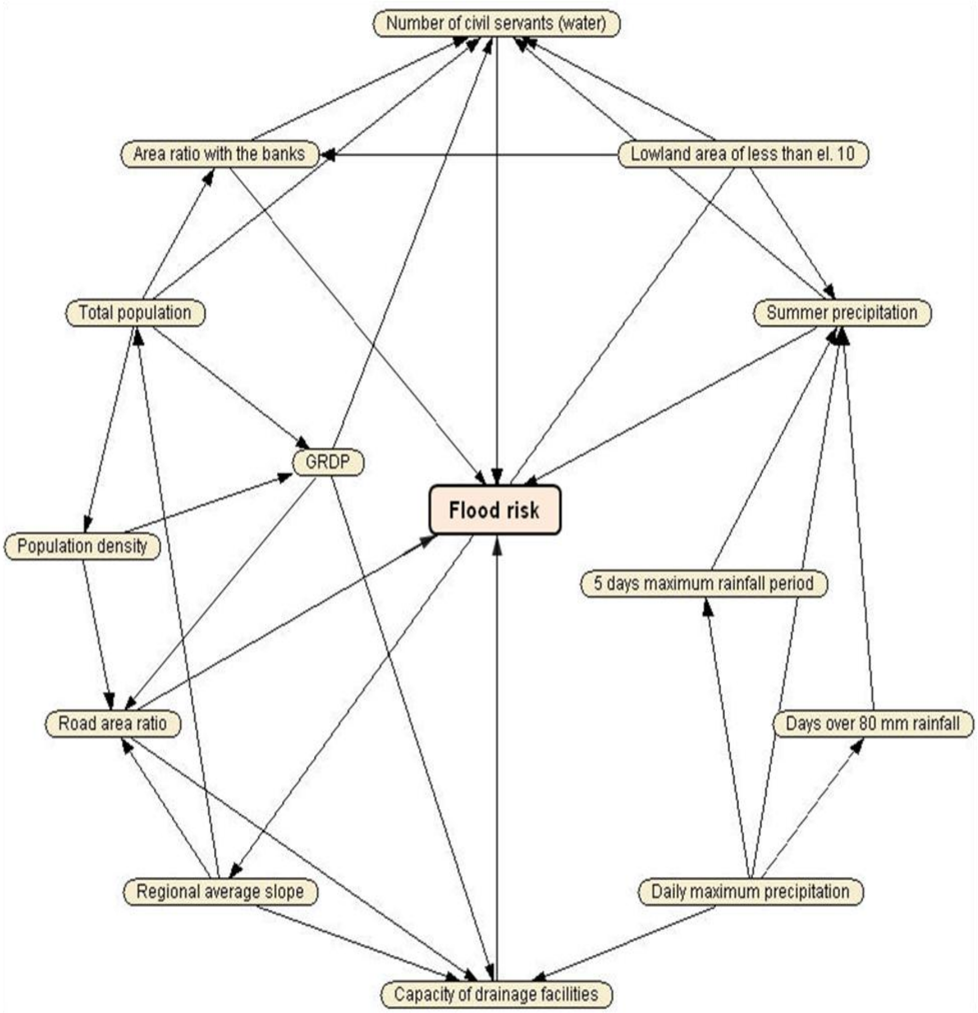
Arcs reflecting the degree of dependencies quantitatively were obtained by calculating CPTs. The iterative process such that establishing arcs and calculating CPTs was carried out. Note that the network was updated repeatedly by changing nodes and direction of arcs as causal relationships among nodes in order to increase validity of the network. Detailed CPTs are attached to Appendix C.

The two Bayesian networks models were established for Seoul and Korea as shown in Figure 3.2 (a) and (b). The causal relationship of nodes is shown in Table 3.7 (a) and (b) for better understanding. In Seoul X_4, X_5, X_6, X_8 and X_{10} (summer precipitation, area ratio with the banks, total population, road area ratio and capacity of drainage facilities) were linked directly to flood risk. In Korea, X_4, X_6, X_8, X_{10} and X_{13} (summer precipitation, total population, road area ratio, capacity of drainage facilities and Number of civil servants related to water) nodes were linked directly to flood.



(a) Seoul

Figure 3.2 Bayesian networks model of flood risk



(b) Korea

Figure 3.2 Bayesian networks model of flood risk (Continued)

Table 3.7 Causal relationship of nodes (a) Seoul

Parent node	Variable	Node	Descendent node
$X_4, X_5, X_6, X_8, X_{10}$	Flood	X_1	
	Daily maximum precipitation	X_2	X_3, X_4, X_{11}
X_2	5 days maximum rainfall period	X_3	X_4
X_2, X_3, X_{11}	Summer precipitation (June to September)	X_4	X_1, X_5, X_{10}
X_4, X_6, X_7, X_{11}	Area ratio with the banks	X_5	X_1
X_7	Total population	X_6	X_1, X_9
	Regional average slope	X_7	X_5, X_6, X_8, X_{10}
X_7, X_9	Road area ratio	X_8	X_1, X_{10}
X_6	Population density	X_9	X_8, X_{10}
$X_4, X_7, X_8, X_9, X_{11}$	Capacity of drainage facilities	X_{10}	X_1
X_2	Days over 80 mm rainfall	X_{11}	X_4, X_{10}

Table 3.7 Causal relationship of nodes (b) Korea

Parent node	Variable	Node	Descendent node
$X_4, X_6,$ X_8, X_{10}, X_{13}	Flood	X_1	
X_2	Daily maximum precipitation	X_2	X_4
X_2	5 days maximum rainfall period	X_3	X_4
X_2, X_3, X_{11}, X_{12}	Summer precipitation (June to September)	X_4	X_1, X_{13}
X_6, X_{12}	Area ratio with the banks	X_5	X_1
X_7	Total population	X_6	X_5, X_9, X_{13}
X_{12}	Regional average slope	X_7	X_6, X_8, X_{10}
X_7, X_9, X_{14}	Road area ratio	X_8	X_1, X_{10}
X_6	Population density	X_9	X_8, X_{14}
X_2, X_7, X_8, X_{14}	Capacity of drainage facilities	X_{10}	X_1
X_2	Days over 80 mm rainfall	X_{11}	X_4
	Lowland area of less than El. 10 m	X_{12}	X_4, X_5, X_7, X_{13}
X_5, X_6, X_{12}, X_{14}	Number of civil servants related to water	X_{13}	X_1
X_6, X_9	GRDP	X_{14}	X_8, X_{10}, X_{13}

4. Model Validation and Application Results

4.1 Flood risk analysis

Comparison with flood risk model and actual flood damage

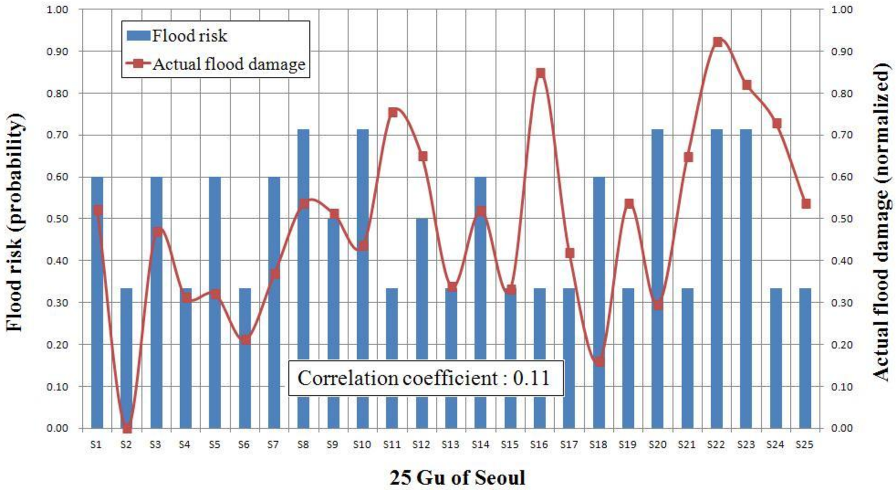
Results of the regional probability of flood risk derived from the BNs model and actual flood damage are shown in Table 4.1 (a) and (b) for Seoul and Korea in increasing order. It reflects probability that a flood might occur or an area might be flooded as affected by incidence of causative factors. Therefore regional potential of flood risk can be compared. Table 4.1 (b) shows probabilities of flood risk of the 13 cities and provinces as regional mean values. Detailed regional probabilities of flood risk of 232 cities • counties • districts across the country are attached to Appendix B. As mentioned above normalized actual flood damage data was used for comparing purpose. To show the reasonable verification of the results, derived probability of flood risks were compared with actual flood damage data taken from Statistics Yearbook for recent 5 years (National Emergency Management Agency, 2006~2010). The compared results of Seoul and Korea are represented in Figures 4.1 (a) and (b).

Table 4.1 Flood risk and actual flood damage (a) Seoul

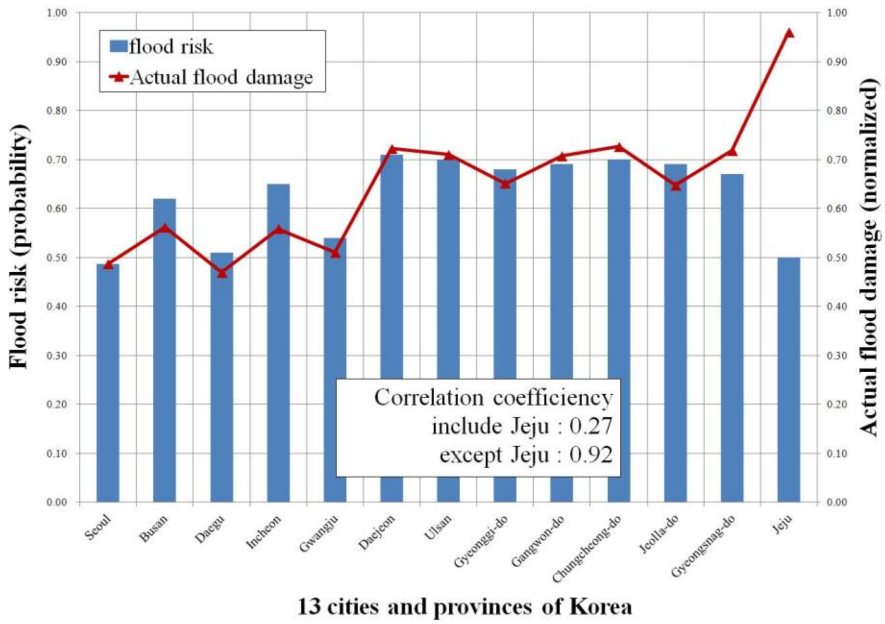
Symbols	Probability of flood risk (%)	Actual flood damage (normalized)	Symbols	Probability of flood risk (%)	Actual flood damage (normalized)
S2	33	0.00	S12	50	0.65
S4	33	0.31	S1	60	0.52
S6	33	0.21	S3	60	0.47
S11	33	0.75	S5	60	0.32
S13	33	0.34	S7	60	0.37
S15	33	0.33	S14	60	0.52
S16	33	0.85	S18	60	0.16
S17	33	0.42	S8	71	0.54
S19	33	0.54	S10	71	0.44
S21	33	0.65	S20	71	0.29
S24	33	0.73	S22	71	0.92
S25	33	0.54	S23	71	0.82
S9	50	0.51			

Table 4.1 Flood risk and actual flood damage (b) Korea

No.	Names of 13 cities and provinces	Probability of flood risk (%)	Actual flood damage (normalized)
1	Seoul	49	0.49
2	Jeju	50	0.98
3	Daegu	51	0.47
4	Gwangju	54	0.51
5	Busan	62	0.56
6	Incheon	65	0.56
7	Gyeongsnag-do	67	0.72
8	Gyeonggi-do	68	0.65
9	Gangwon-do	69	0.71
10	Jeolla-do	69	0.65
11	Ulsan	70	0.71
12	Chungcheong-do	70	0.73
13	Daejeon	71	0.72



(a) Seoul



(b) Korea

Figure 4.1 Comparison of flood risk

A quantitative comparison between results of flood risk and actual flood damage could be achieved by calculation of the correlation coefficient. In Seoul, correlation coefficient was 0.11. Especially, S11 (Nowon-Gu), S16 (Gangseo-Gu), S18 (Geumcheon-Gu), S20 (Dongjak-Gu) showed different tendency with actual flood damage. Even though the value of correlation coefficient was very low, the average of flood risk (49 %) from Bayesian networks and the normalized value of actual flood damage (0.49) indicated some importance. It means general result of flood risk obtained from Bayesian networks has validity. The correlation coefficient of Korea was 0.27 but if excluding Jeju it was 0.92. Jeju consist of just 2 cities so data used to calculate flood risk was very few cases. While the normalized value of actual flood damage of Jeju recorded the highest in Korea as 0.99 and 1. 2 sub-districts were hard to calculate overall results of flood risk in Jeju. The numbers of case have strong influence to accuracy of results of flood risk. Excluding Jeju, flood risk model for Korea indicated an excellent trend. Therefore flood risk derived Bayesian networks had high accuracy and validity to actual flood damage. The results of verification provided importance of considering causal relationships among variables and meaningful conditional distributions.

Correlation coefficient between flood risk and actual flood damage was used to verify effectiveness of Bayesian networks for flood risk. For detailed analysis the correlation values of Korea were divided into 3 levels as 0～0.3, 0.3～0.5 and 0.5 ～1. (Table 4.2) Chungcheong-do showed the lowest result as 0.077, Daejeon had the highest result as 0.802. High correlation group (Busan, Daegu, Gwangju, Daejeon) showed good performance about adaptive capacity like capacity of drainage facilities contributing to reduce the flood damage although they were highly sensitive to impact of climate exposure like high total population and population density, road area ratio and low climate exposure like daily maximum precipitation, days over 80 mm rainfall and 5 days maximum rainfall period. On the other hand, a group of low correlation coefficient indicated 0～0.3 including Chungcheong-do, Seoul, Jeolla-do and Gyeongsang-do recorded high adaptive capacity and high climate exposure. Thus, unpredictable climate exposure is the most influence factor in comparing result with actual flood damage. Adaptive capacity to reduce the climate change can be contribute to predict flood risk because it was evaluated with more accuracy.

Table 4.2 Correlation coefficients of flood risk model

No.	Correlation coefficient		Cities and provinces
	Range	Value	
1	0 ~ 0.3	0.077	Chungcheong-do
2		0.109	Seoul
3		0.172	Jeolla-do
4		0.177	Gyeongsang-do
5	0.3 ~ 0.5	0.369	Gangwon-do
6		0.380	Gyeonggi-do
7		0.388	Ulsan
8		0.485	Incheon
9	0.5 ~ 1.0	0.503	Busan
10		0.574	Daegu
11		0.614	Gwangju
12		0.802	Daejeon

Sensitivity analysis of proxy variables to flood risk

CPTs and ranges of each nodal value were inputted considering relationships of connected nodes by arcs. The network was compiled and sensitivities of proxy variables to flood risk were obtained by Netica^{TM}.

Mutual information (Shannon and Weaver 1949) is one of the most commonly used measures for ranking information sources. It is based on the assumption that the uncertainty regarding and variable Z characterized by a probability distribution $\Pr(Z)$ can be represented by the entropy function

$$H(T) = -\sum_{z} \Pr(T) \log \Pr(T) \tag{4.2}$$

Accordingly, the residual uncertainty regarding the true value of the target variable T (flood risk in this research), given that Y, can be written

$$H(T|Y) = -\sum_{t} \Pr(T|Y) \log \Pr(T|Y) \tag{4.3}$$

And the average residual uncertainty in T, summed over all possible outcomes Y, is

$$H(T|Y) = \sum_y H(T|Y)\Pr(Y)$$
$$= -\sum_y \sum_t \Pr(T,Y)\log \Pr(T|Y) \tag{4.4}$$

If we subtract $H(T|Y)$ from the original uncertainty in T prior to consulting Y, namely $H(T)$, we will obtain the total uncertainty-reducing potential of Y. This potential is called Shannon's mutual information and is given by

$$I(T,Y) = H(T) - H(T|Y)$$
$$= -\sum_y \sum_t \Pr(T,Y)\log \frac{\Pr(T,Y)}{\Pr(T)\Pr(Y)} \tag{4.5}$$

Clearly, $I(T,Y)$ is symmetric with respect to Y and T, so

$$I(T,Y) = I(Y,T) = H(Y) - H(Y|T) \tag{4.6}$$

Additionally, $I(T,Y)$ is a nonnegative quantity and is equal to 0 if and only if T and Y are mutually independent (Pearl, 1988).

The relative influence of each proxy variable on the flood risk is shown in Table 4.3 (a) and (b) in decreasing order.

Table 4.3 Sensitivity analysis results in decreasing order of influence on flood

(a) Seoul

Node (variable)	Sensitivity (%)
Flood	100
Road area ratio	0.46
Total population	0.38
Summer precipitation (June to September)	0.307
Area ratio with the banks	0.18
Regional average slope	0.14
Capacity of drainage facilities	0.0258
5 days maximum rainfall period	0.0073
Daily maximum precipitation	0.0069
Days over 80 mm rainfall	0.0045
Population density	0.0031

Table 4.3 Sensitivity analysis results in decreasing order of influence on flood

(b) Korea

Node (variable)	Sensitivity (%)
Flood	100
Number of civil servants related to water	5.06
Area ratio with the banks	3.01
Summer precipitation (June to September)	1.24
Capacity of drainage facilities	1.04
Road area ratio	0.827
GRDP	0.118
Lowland area of less than El. 10 m	0.0952
Total population	0.0414
Days over 80 mm rainfall	0.0317
Daily maximum precipitation	0.0263
5 days maximum rainfall period	0.0212
Regional average slope	0.0157
Population density	0.0026

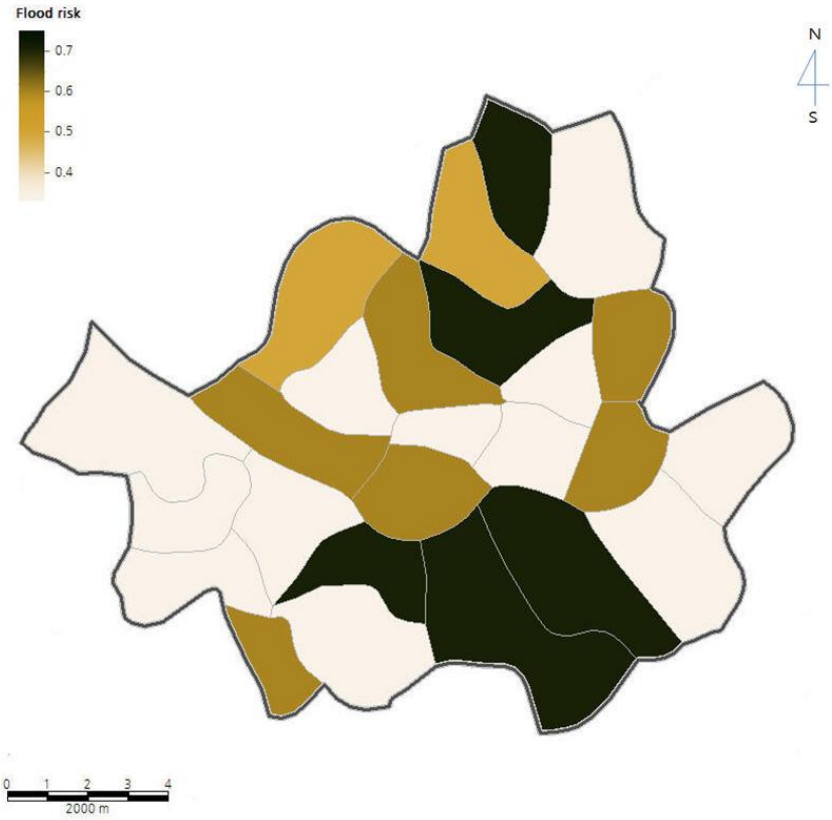
The influence of a specific proxy variable to the flood risk was represented through sensitivity analysis. In Bayesian networks model for Seoul the biggest sensitive proxy variable to flood risk was road area ratio as 0.46 %. The second variable was total population as 0.38 % and summer precipitation was 0.307 %. Area ratio with the banks and regional average slope showed sensitivities of 0.18 and 0.14 %. The remainders had sensitivity less than 0.2 %.

This indicates that the characteristic of urbanization and large population in Seoul have an effect. In case of Korea, the biggest was number of civil servants related to water as 5.06 % and the second was area ratio with the bank as 3.01 %. Summer precipitation and capacity of drainage facilities showed sensitivities of 1.24 % and 1.04 %. Road area ratio and GRDP showed 0.827 and 0.118 %. The remainders had sensitivities less than 1 %. Adaptive capacity to reduce the flood risk such as number of civil servants and capacity of drainage facilities denoted strong influence. Sensitivities of the network for Seoul were obtained as very small. It is supposed that amount of data is not sufficient for Seoul because number of sub-districts in Seoul is 25, just 1/9 of Korea. It can be inferred that amount of data for each node is very important to flood risk analysis.

4.2 Flood risk diagnosis

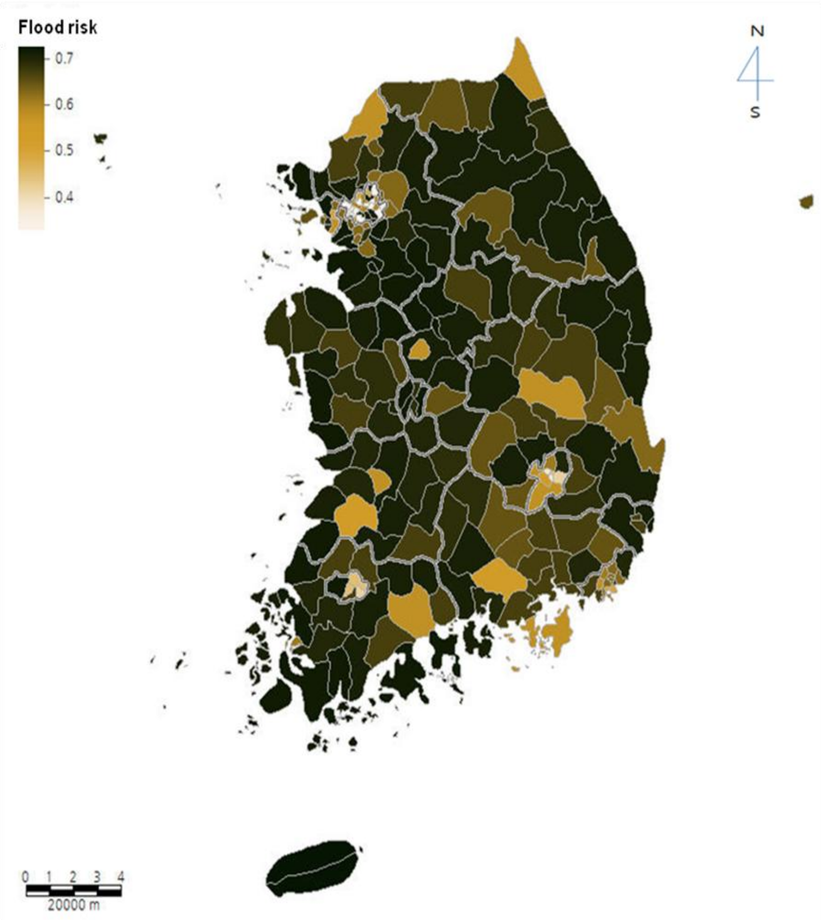
The flood risk mapping was presented based on the results of probability of flood risk to diagnose flood risk using CCGIS^{TM}. The spatial distributions of the flood risk results for Seoul and Korea are shown in Figures 4.2 (a) and (b). For comparison, the actual flood damage of Seoul and Korea was shown in Figures 4.3 (a) and (b).

The mean value of flood risk of Korea is 63%, Seoul, Daegu, Gwanju and Busan were below the average in Table 4.1 (b). The rests are higher than average. Dajeon, Chungcheong-do and Ulsan showed higher flood risk than other area. Their number of civil servants, area ratio with the banks and capacity of drainage facilities were higher than Seoul, Daegu, Gwanju relatively. That is the higher sensitivity to flood risk, the higher flood risk. The sensitivity of the proxy variable should consider for flood risk management.



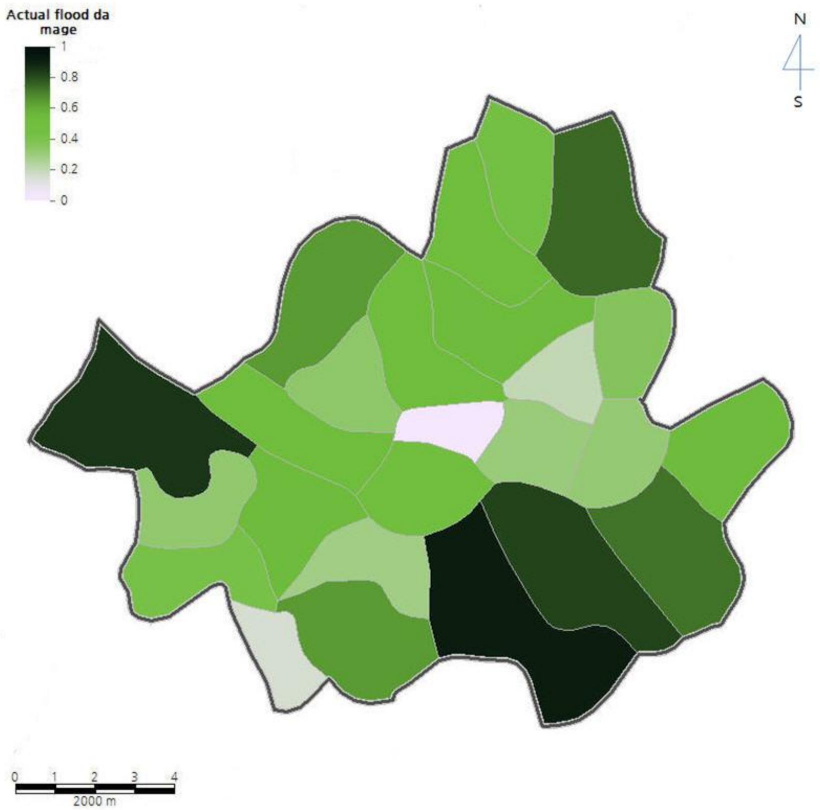
(a) Seoul

Figure 4.2 Flood risk map derived from Bayesian networks model



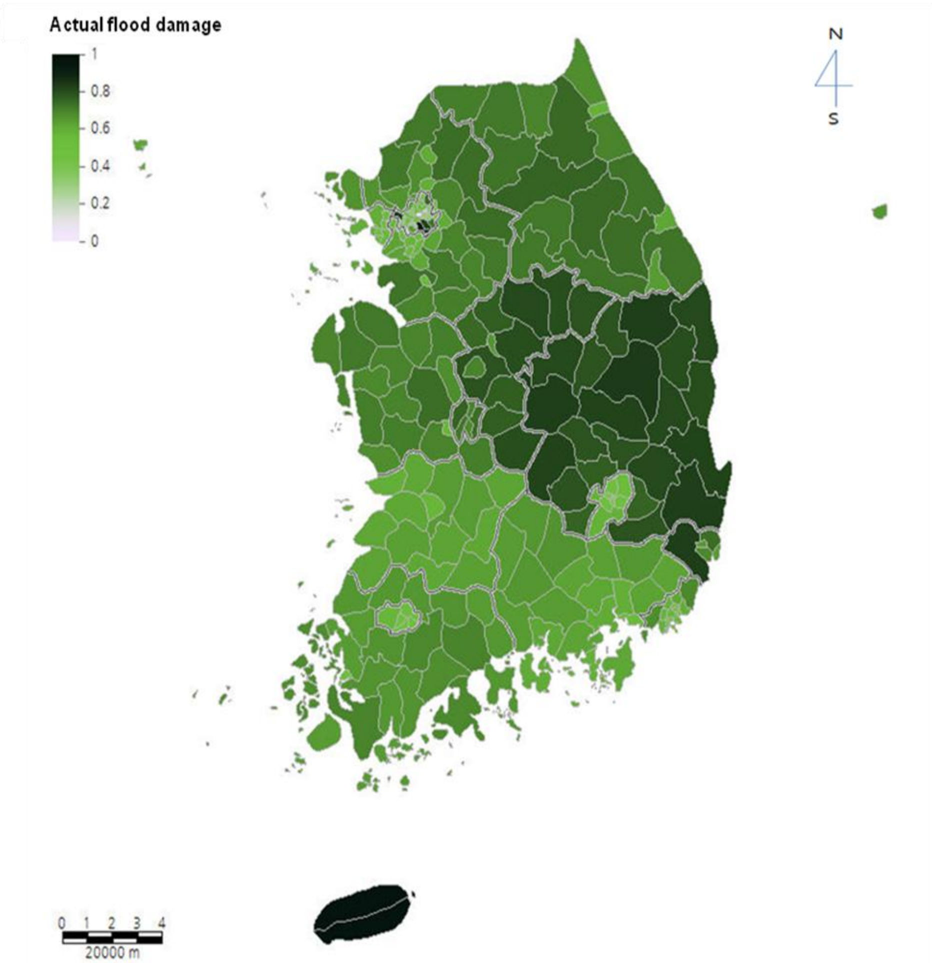
(b) Korea

Figure 4.2 Flood risk map derived from Bayesian networks model (Continued)



(a) Seoul

Figure 4.3 Flood risk map derived from the actual flood damage



(b) Korea

Figure 4.3 Flood risk map derived from the actual flood damage (Continued)

5. Conclusions and Future Study

5.1 Conclusions

This research is the first attempted assessment of flood risk using Bayesian networks in Korea. BNs approach was used to derive probabilistic flood risk maps using CCGIS™ and raw data through integration of proxy variables. Despite the limitation of the data used, the results reveal the practical usefulness of the innovative method suggested in this research. Proposed flood risk network can be composed of actual factors related to flood risk. Through the network model we can notice casual relationships among proxy variables related to flood risk. In Seoul, proxy variables of road area ratio, total population, summer precipitation had sensitivity to flood risk. This reflected that the characteristic of urbanization and large population in Seoul. In Korea, number of civil servants related to water, area ratio with the banks, summer precipitation, and capacity of drainage facilities showed significant sensitivity to flood risk. And especially flood risk of big cities such as Incheon, Busan, Daegu, Gwangju and Daejeon showed higher accuracy with actual flood damage. The analysis results mean significance of adaptation capacity to reduce the impact of climate change. It can evaluate flood risk quantitatively.

The proposed flood risk model showed high accuracy and validity to actual flood damage. We can guess the reason that proxy variable data was based on actual data. It shows that the flood risk model can predict future flood risk in Korea. Furthermore it can indicate the direction of future plan for flood damage reduction where to go and give us an indication about variables that affect flood risk probability and how the probability can be reduced through flood risk management. Nowadays public interest in respect of flood and drought and importance of social recognition and agreement of water control policy increase due to impact of climate change. The flood risk map derived in this research will be of great help to flood control managers and political decision makers for flood risk assessment and mitigation planning as useful information.

5.2 Future study

Further work is needed to improve development of proxy variables and data investigation for practical and specific applications in flood risk analysis. Also application of Bayesian networks in other hydrologic model can be suggested to identify the uncertainty. Water control, water use and water quality are important in integrated water resources management. Therefore in future research of flood risk consideration of these factors is positively necessary. Finally, plan for political decision making for flood risk management reflecting this result should be studied.

References

Bernier, J. (1994a). “Quantitative analysis of uncertainties in water resources: application for predicting the effects of changes.” *Engineering Risk in Natural Resources Management*, Vol. 275, pp. 343-358.

Bernier, J. (1994b). “Statistical detection of changes in geophysical series.” In *Engineering Risk and Reliability in a Physical Environment*, Vol. 275, pp. 159-176.

Chung, E. S. and Lee, K. S. (2007). “Identifying spatial hazard ranking using multicriteria decision making techniques.” *Journal of Korean Water Resources Association.*” Vol. 40, No. 12, pp. 969~983.

Daniel, B. K., Zapata-Rivera, J.-D., and McCalla, G. I. (2007). “A Bayesian belief network approach for modeling complex domains.” *Bayesian Network Technologies: Applications and Graphical* (Mittal, A., Kassim, A., and Tan, T. eds.)., pp. 13–51. IGI.

Dlamini, W. M. (2011). “Application of Bayesian networks for fire risk mapping using GIS and remote sensing data.” *Geo Journal*, Vol. 76, No. 3, pp. 283-296.

Duckstein, L., Bobee, B., and Bogardi, I. (1987). “Bayesian forecasting of hydrologic variables under changing climatology.” In *The Influence of Climate Change and Climatic Variability on the Hydrologic Regime and Water Resources*, Vol. 168, pp, 301-311.

Filliben, T. T. (1975). “The probability plot correlation coefficient test for normality.” *Technometrics*, Vol. 17, No. 1, pp. 111-117.

Han, K. Y., Kim, K. S., and Lee, Y. S. (2007). “Establishment methodologies for flood information map.” *Proceedings of Korean Water Resources Association Conference* (in Korean).

Hatfield Consultants Ltd. (2006). *Using participatory methodologies, geographic information systems and earth observation data to map traditional ecological knowledge in Hong Ha Commune, Thua Thien Hue, Viet Nam. Milestone 9 Report, EOSTEM Project*. Hatfield Consultants, West Vancouver, Canada.

Heckerman, D. (1995). “A Bayesian approach for learning causal networks.” *Proceedings of Eleventh Conference on Uncertainty in Artificial Intelligence*. Montreal, QU.

Howard, R. A. and Matheson, J. (1983). *The principles and applications of decision analysis*. Strategic Decisions Group, Menlo Park, CA.

IPCC (2012). *Managing the risks of extreme events and disasters to advance climate change adaptation*. Cambridge University Press.

Jang, O. J. and Kim, Y. O. (2009). “Flood risk estimation using regional regression analysis.” *Journal of Korean Society of Hazard Mitigation*, Vol. 9, No. 4, pp. 71-80 (in Korean).

Kannami, Y. (2008). “Establishment of country-based flood risk index.” *Proceedings of 30th ICHARM Internationl Syposium Conference.*

Karmakar, A.P.S. and Simonovic, S.P. (2007). *Physical, economical, infrastructural and social flood risk – Vulnerability analyses in GIS*. Water Resources Research Report, Report 57, No. 1913-3200, The University of Western Ontario, Department of Civil and Environmental Engineering, Canada.

Kim, S. G., Song, I. H., and Ko, K. Y. (2007). “Cartography of disaster in Yeongsan River watershed using geographical information” *Journal of Korea Cadastre Information Association*, Vol. 9, No. 2, pp. 121-131 (in Korean).

Kim, S. U. (2007). *Low flow frequency analysis using Bayesian approach*. Ph. D. Thesis, Seoul National University, Korea.

Kim, T. H., Han, K. Y., and Cho, W. H. (2011). “Vulnerability analysis in the Nakdong River basin for the utilization of flood risk mapping.” *Journal of Korean Association of Geographic Information Studies*, Vol. 68, No. 3, pp. 203-222 (in Korean).

Kottegoda, N.T. and Rosso, R. (1997). *Statistics, probability, and reliability for civil and environmental engineer*. McGraw-Hill, New York, NY.

Lee, C. H., Lee, S. M., and Yeo, C. G. (2006). *Development of the regional safety assessment model in Seoul - Focusing on flood -* . Seoul Development Institute, No. 1 (in Korean).

Lim, K. S., Choi, S. J., Lee, D. R., and Moon, J. W. (2010). “Development of flood risk index using causal relationships of flood indicators” *Journal of Korean Society of Civil Engineers*, Vol. 30, No. 1B, pp. 61-70 (in Korean).

Marcot, B. G., Steventon, J. D., Sutherland, G. D., and McCann, R. K. (2006). “Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation.” *Canadian Journal of Forestry Research*, Vol. 36, No, 12, pp. 3063-3074.

McCann, R. K., Marcot, B. G., and Ellis, R. (2006). “Bayesian belief networks: Applications in ecology and natural resource management.” *Canadian Journal of Forestry Research*, Vol. 36, No. 12, pp. 3053–3062.

Ministry of Land, Transport and Maritime Affairs (2006). *Research of basic strategy setting for water resources informatization* (in Korean).

National Emergency Management Agency (2006~2010). *Disaster yearbook* (in Korean).

National Institute of Environmental Research (2012). *CCGIS ver 3.1 manual* (in Korean).

National Institute of Environmental Research (2012). *Climate change department vulnerability mapping* (in Korean).

Noson, L. (2002) "Hazard mapping and risk assessment." *Proceedings of Regional Workshop on Best Practices in Disaster Mitigation*.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann Publisher, San Francisco.

Pollino, C. A., Woodberry, O., Nicholson, A., Korb, K., and Hart, B. T. (2007). "Parameterisation and evaluation of a Bayesian belief network for use in an ecological risk assessment." *Environmental Modelling & Software*, Vol. 22, No. 8, pp. 1140–1152.

Shannon, C. and Weaver, W. (1949). *The mathematical theory of communication*, University of Illinois Press.

Spiegelhalter, D. J., Dawid, A. D., Lauritzen, S. L., and Cowell, R. G. (1993). "Bayesian analysis in expert systems." *Statistical Science*, Vol. 8, No. 3, pp. 219–283.

Tukey, J. (1977). *Exploratory data analysis*. Addison, Wesley.

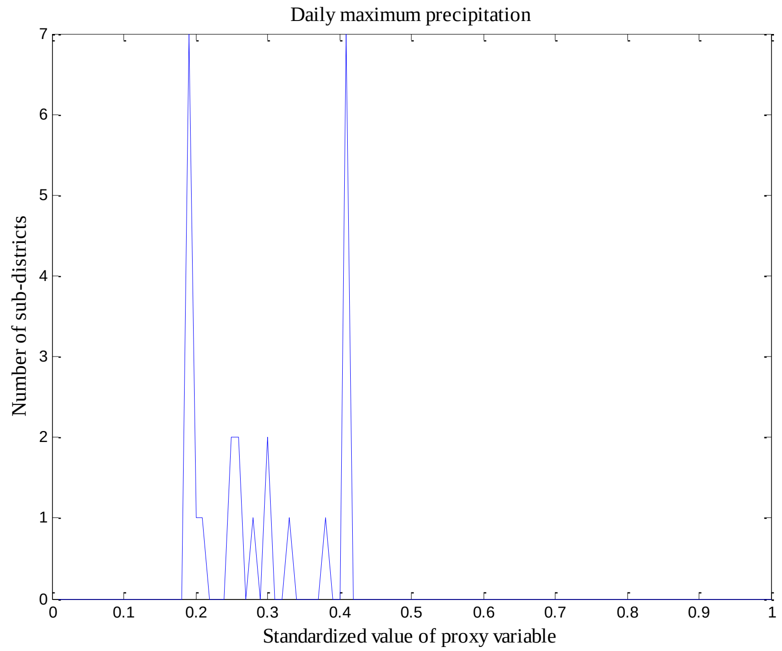
Uusitalo, L. (2007). "Advantages and challenges of Bayesian networks in environmental modeling." *Ecological Modelling*, Vol. 203, No. 3–4, pp. 312–318.

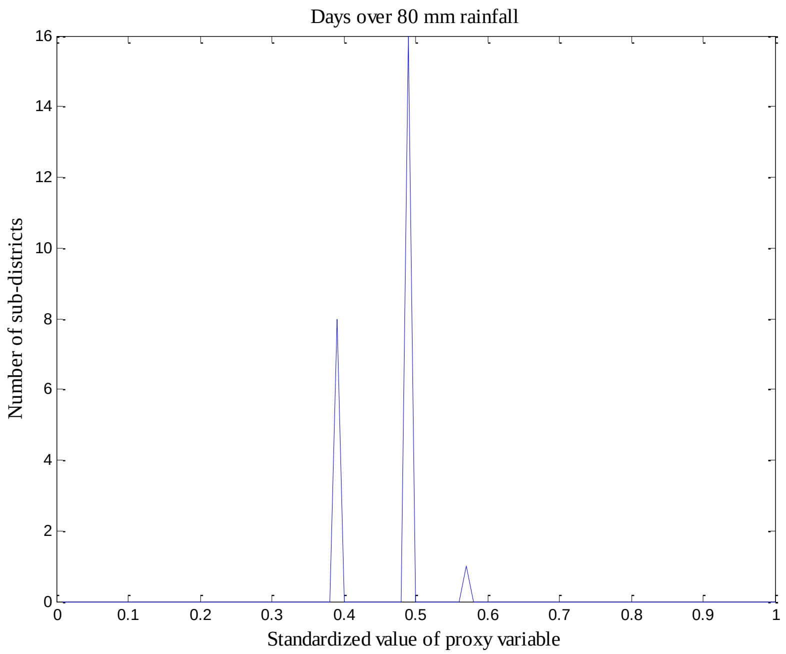
Zhang, J., Okada, N., Tatano, H., and Hayakawa, S. (2002). "Risk assessment and zoning of flood damage caused by heavy rainfall in Yamaguchi Prefecture, Japan." *Flood Defence 2002*. Science Press, New York.

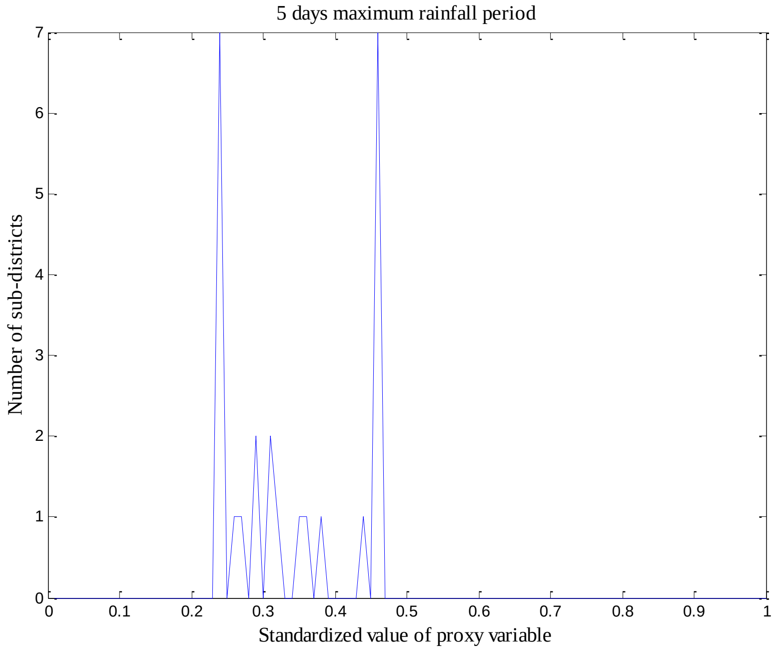
Korea Meteorological Administration: www.kma.go.kr

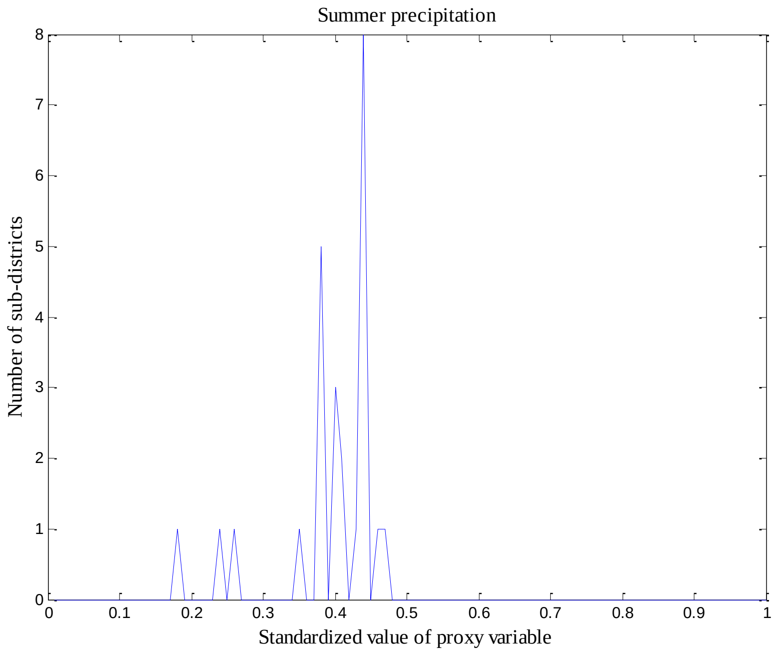
Norsys Software Corporation: www.norsys.com

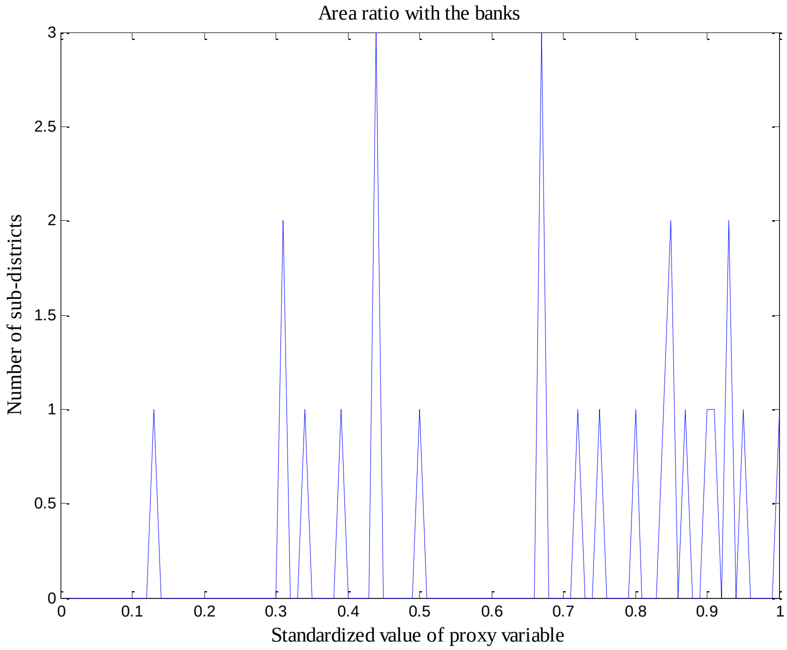
Appendix A1. Histograms of proxy variables for Seoul

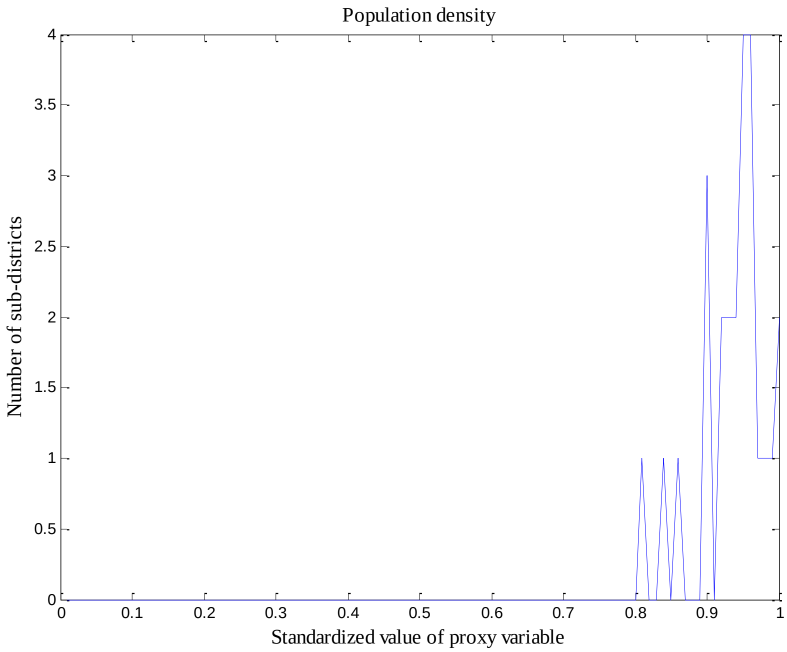


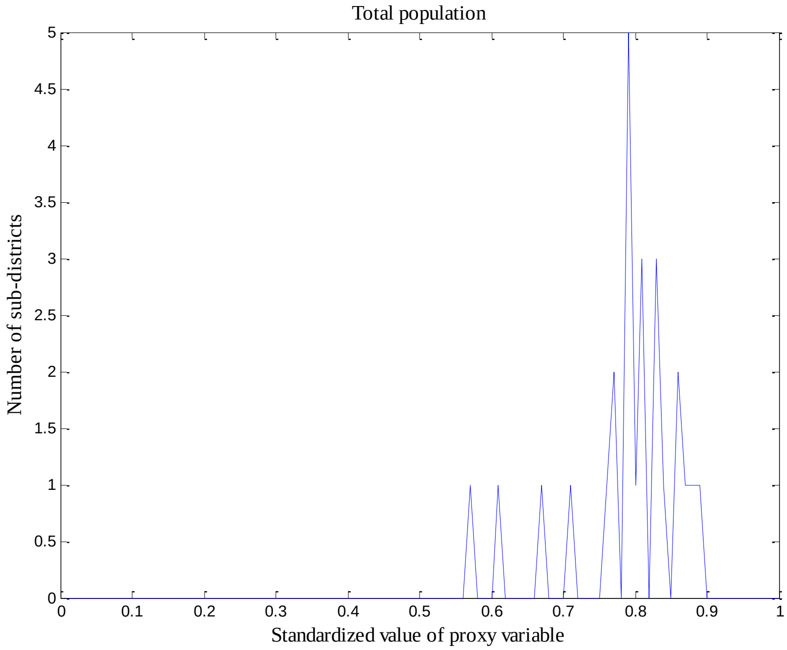


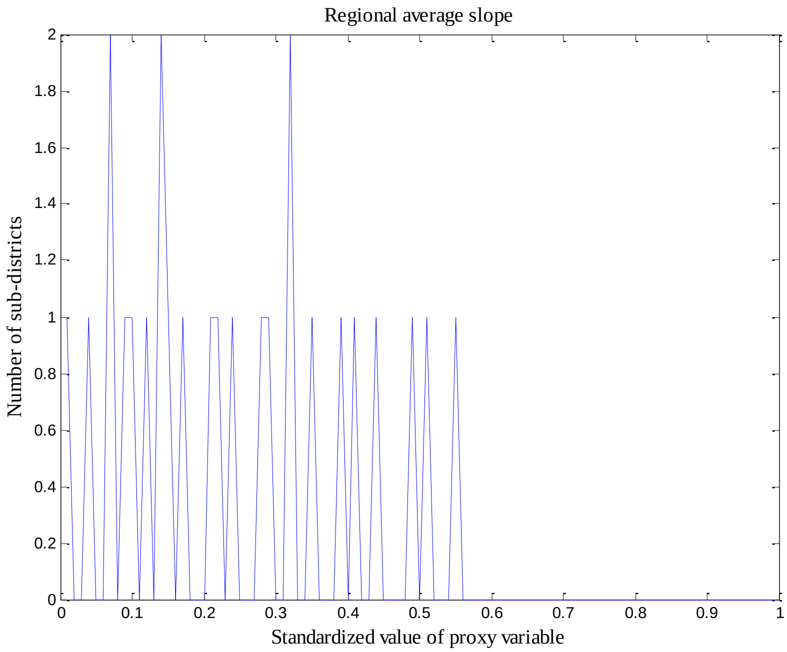

5 days maximum rainfall period
Number of sub-districts
Standardized value of proxy variable
0
1
2
3
4
5
6
7
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

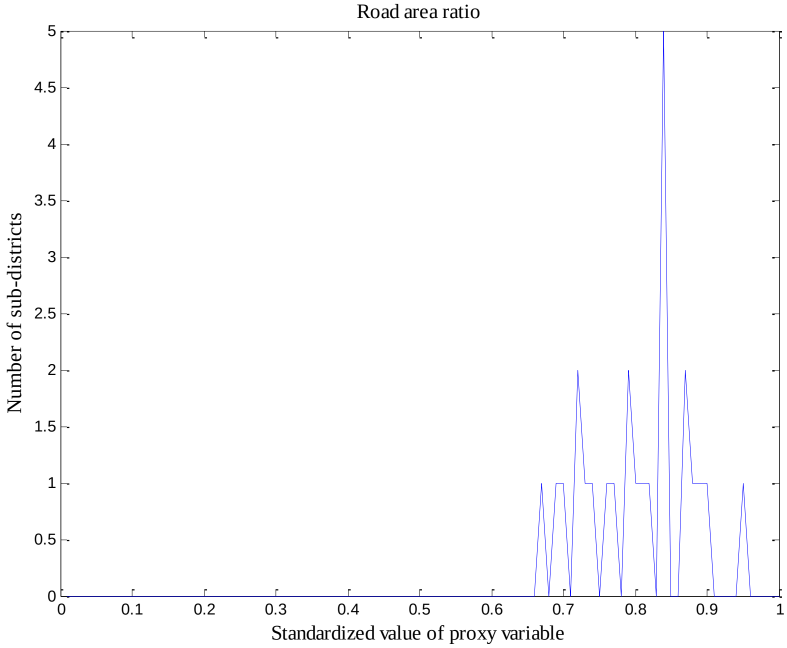

Summer precipitation
Number of sub-districts
Standardized value of proxy variable
0
1
2
3
4
5
6
7
8
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

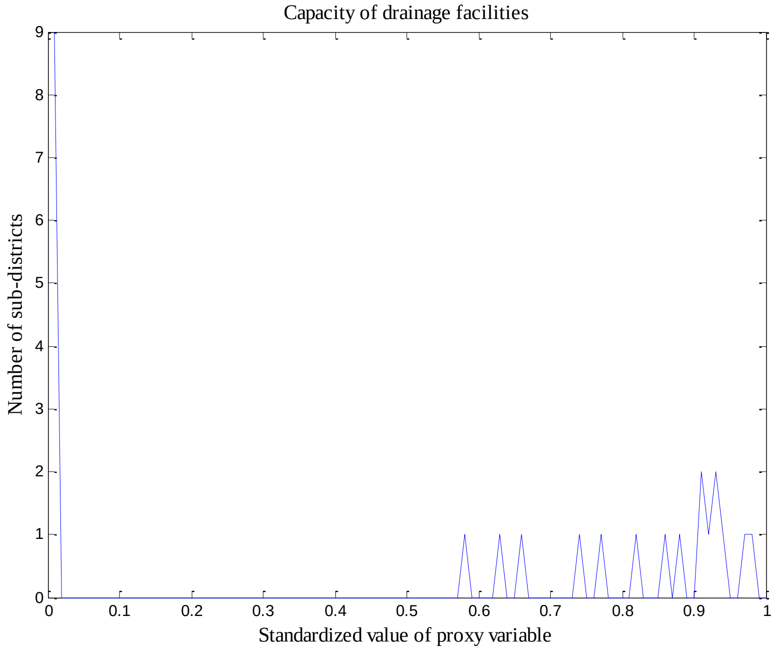

Area ratio with the banks
Number of sub-districts
Standardized value of proxy variable
0
0.5
1
1.5
2
2.5
3
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


Population density
Number of sub-districts
Standardized value of proxy variable
0
0.5
1
1.5
2
2.5
3
3.5
4
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

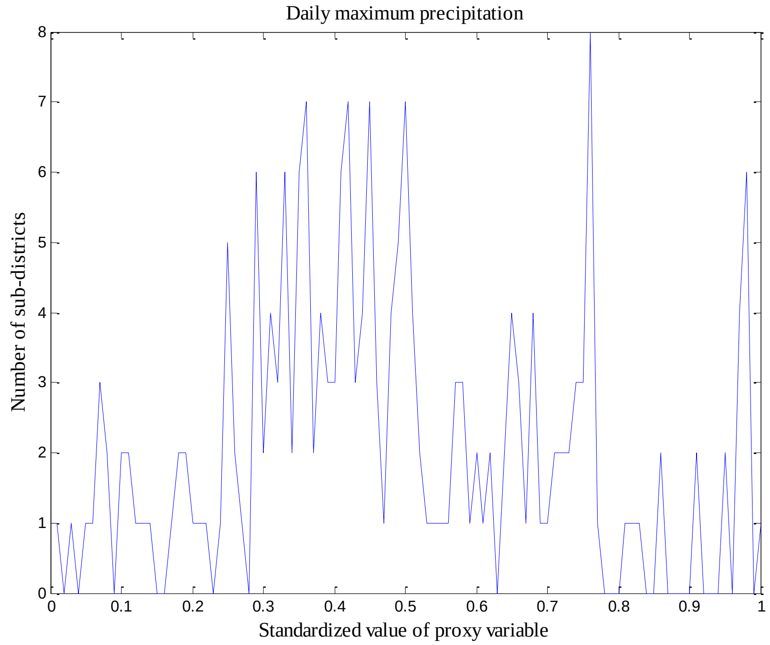

Total population
Number of sub-districts
Standardized value of proxy variable
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

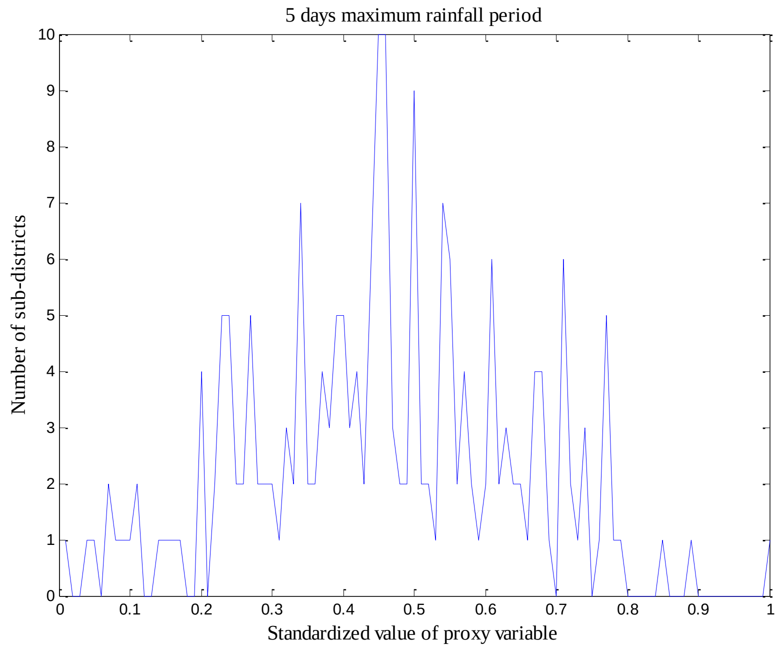

Regional average slope
Number of sub-districts
Standardized value of proxy variable
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

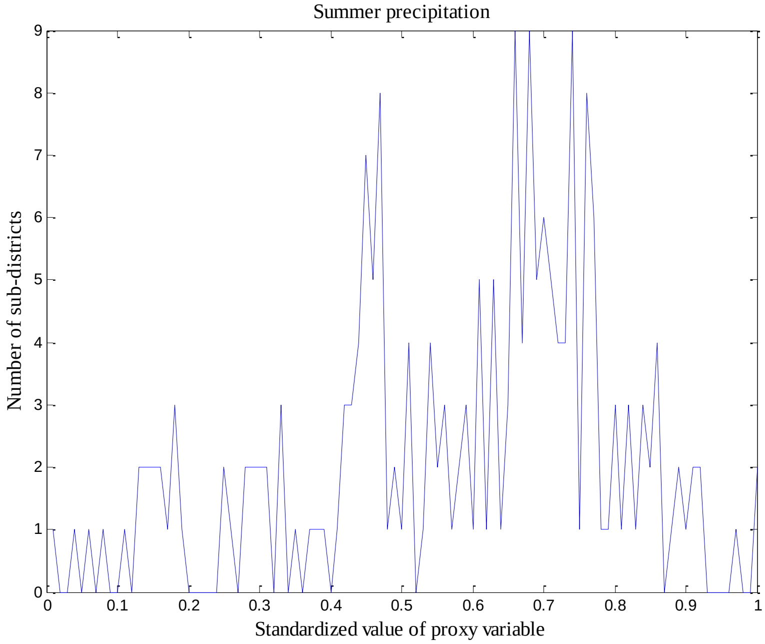

Road area ratio
Number of sub-districts
Standardized value of proxy variable
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

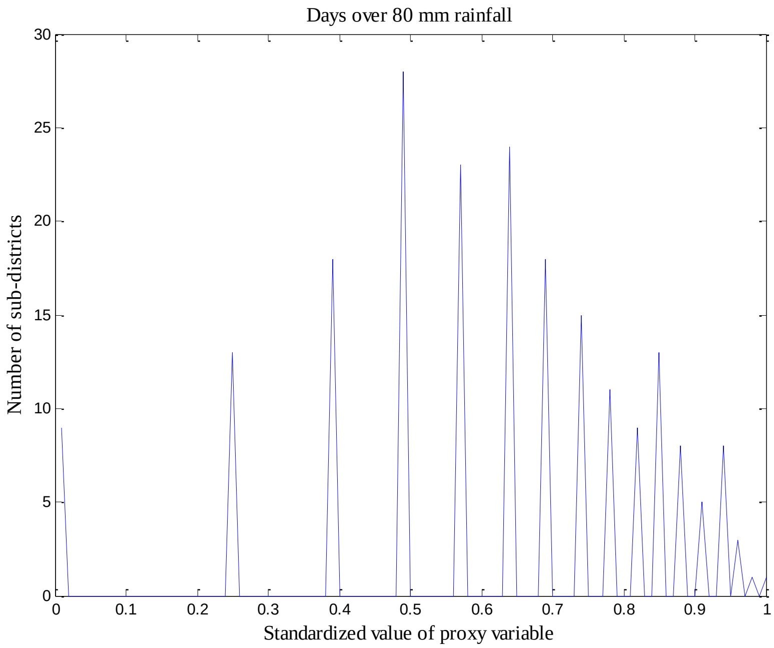

Capacity of drainage facilities
Number of sub-districts
Standardized value of proxy variable
0
1
2
3
4
5
6
7
8
9
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

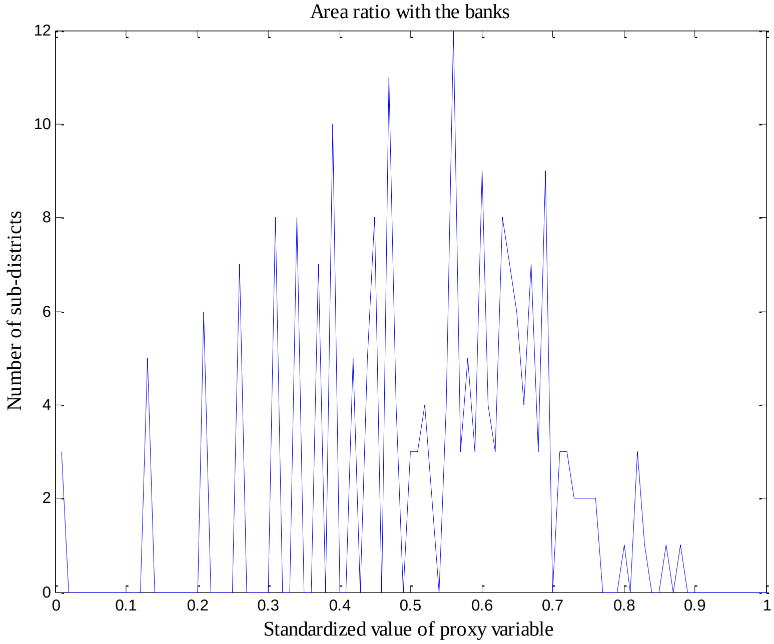
Appendix A2. Histograms of proxy variables for Korea

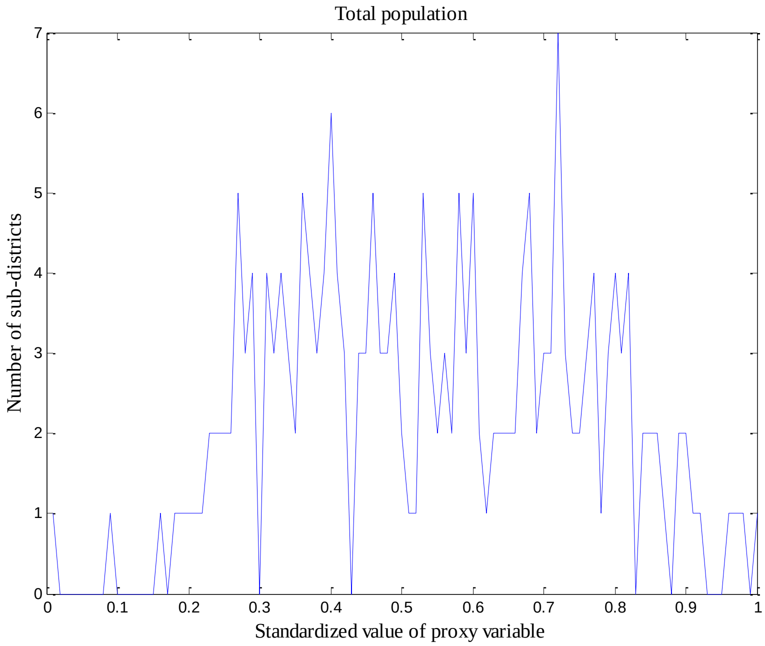


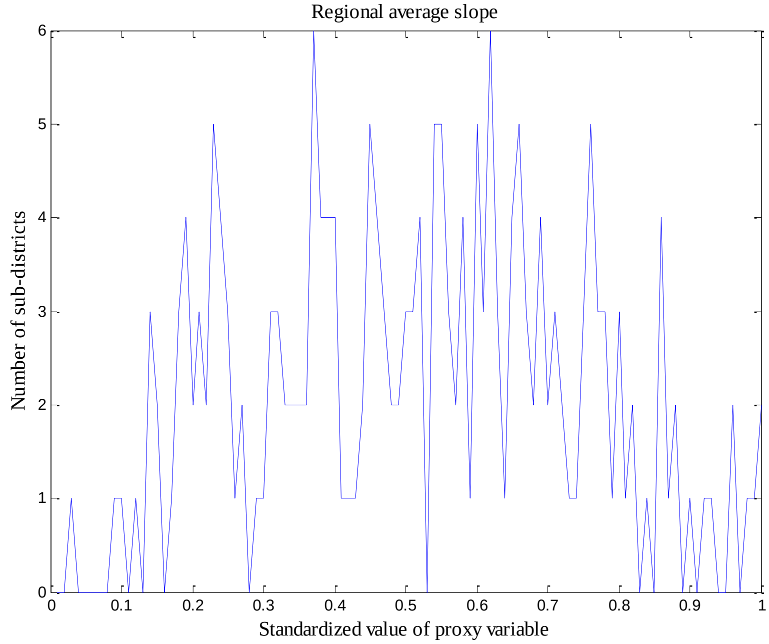


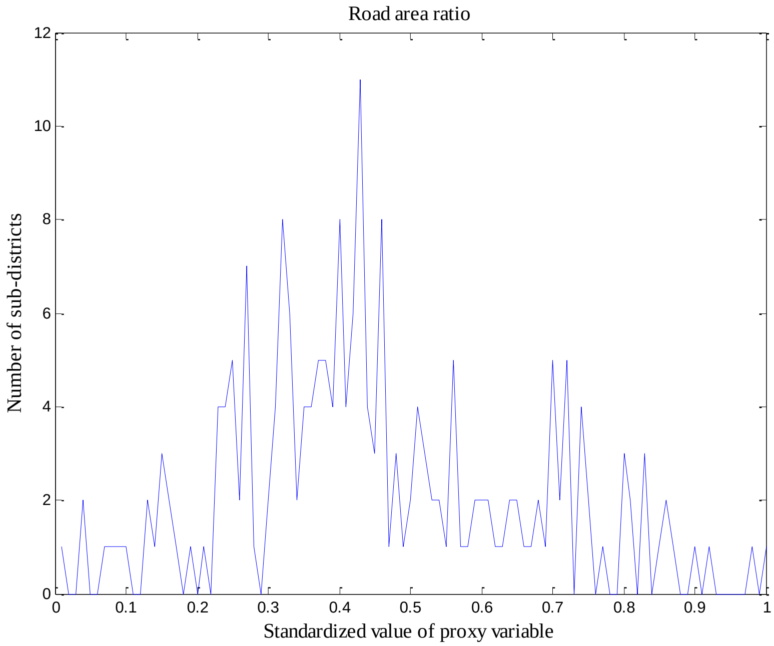

Summer precipitation
Number of sub-districts
Standardized value of proxy variable
0
1
2
3
4
5
6
7
8
9
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


Days over 80 mm rainfall
Number of sub-districts
Standardized value of proxy variable
0
5
10
15
20
25
30
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

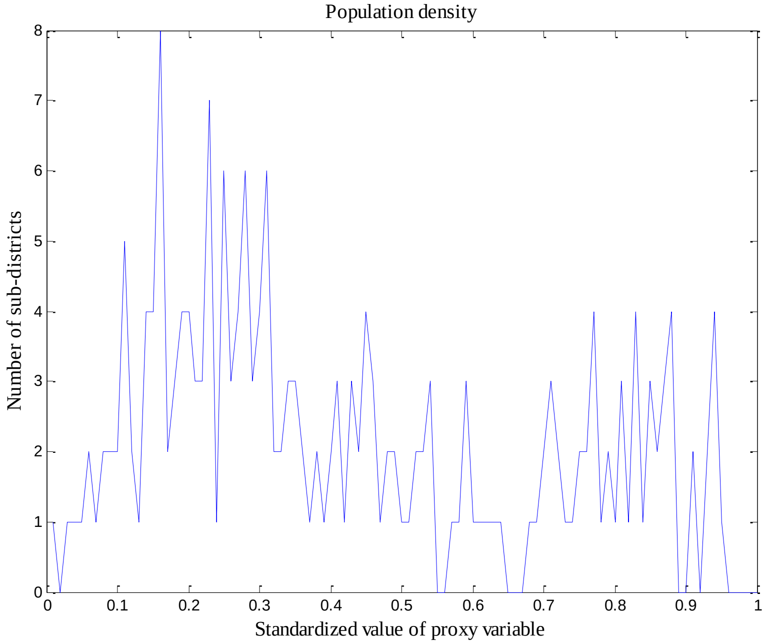

Area ratio with the banks
Number of sub-districts
Standardized value of proxy variable
0
2
4
6
8
10
12
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


Total population
Number of sub-districts
Standardized value of proxy variable
0
1
2
3
4
5
6
7
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

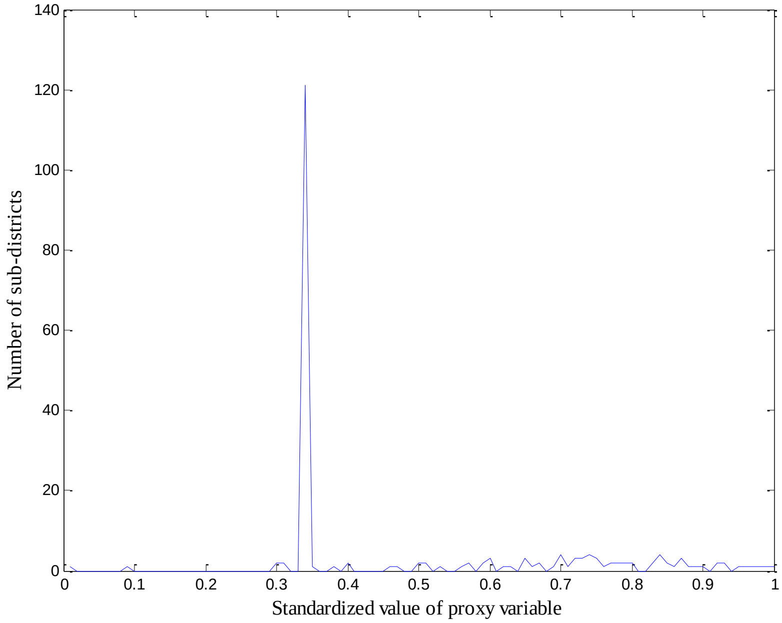

Regional average slope
Number of sub-districts
Standardized value of proxy variable
0
1
2
3
4
5
6
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

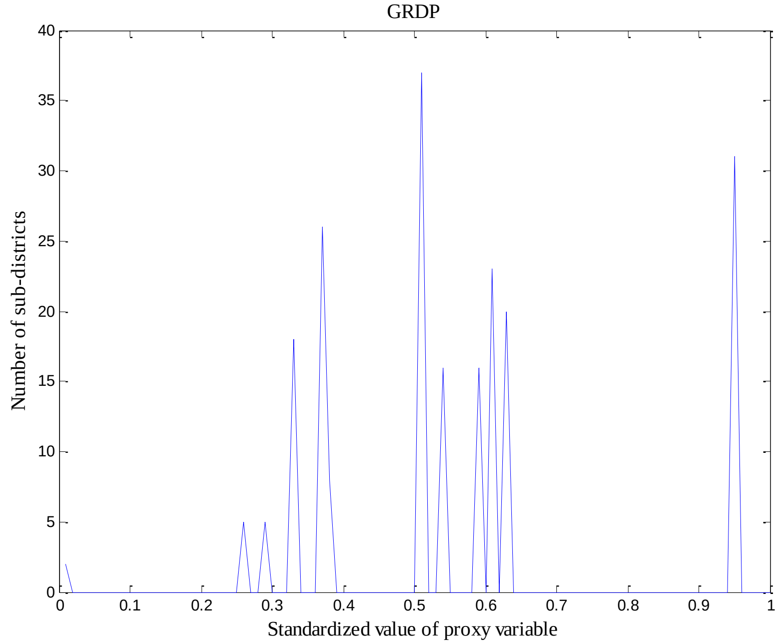

Road area ratio
Number of sub-districts
Standardized value of proxy variable
0
2
4
6
8
10
12
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

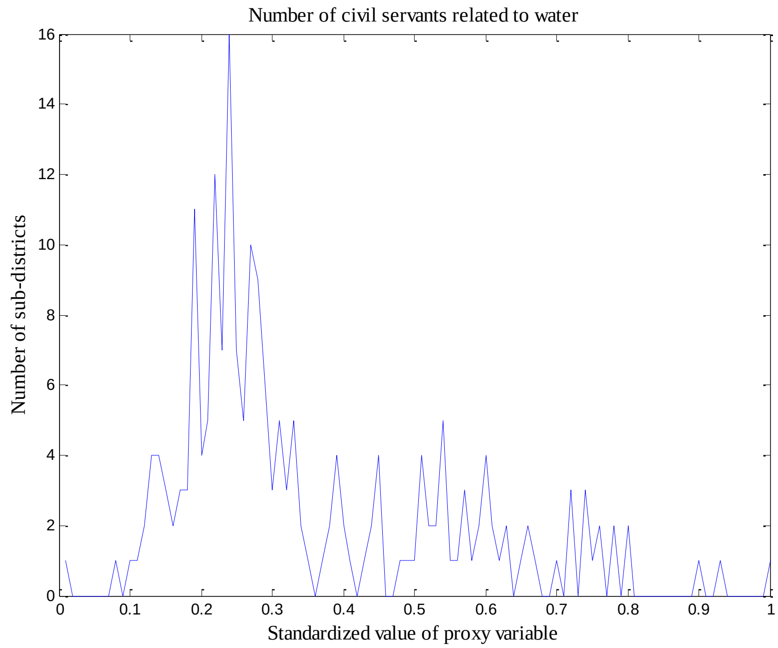
Population density

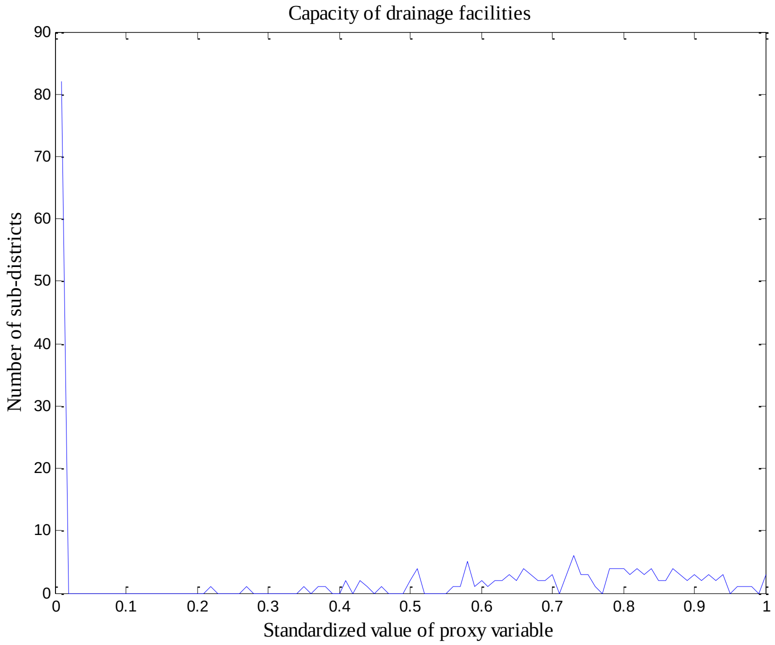


Lowland area of less than el. 10m




GRDP
Number of sub-districts
Standardized value of proxy variable
0
5
10
15
20
25
30
35
40
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


Number of civil servants related to water
Number of sub-districts
Standardized value of proxy variable
0
2
4
6
8
10
12
14
16
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


Capacity of drainage facilities
Number of sub-districts
Standardized value of proxy variable
0
10
20
30
40
50
60
70
80
90
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Appendix B. Probability of flood risk using the Bayesian networks model for 13 cities and provinces

No.	Local Governments and sub-districts		Probability of flood risk (%)
1	Seoul	S1	60.00
2		S2	33.40
3		S3	60.00
4		S4	33.40
5		S5	60.00
6		S6	33.40
7		S7	60.00
8		S8	71.40
9		S9	50.00
10		S10	71.40
11		S11	33.40
12		S12	50.00
13		S13	33.40
14		S14	60.00
15		S15	33.40

No.	Local Governments and sub-districts		Probability of flood risk (%)
16	Seoul	S16	33.40
17		S17	33.40
18		S18	60.00
19		S19	33.40
20		S20	71.40
21		S21	33.40
22		S22	71.40
23		S23	71.40
24		S24	33.40
25		S25	33.40
26	Busan	B1	75.00
27		B2	75.00
28		B3	50.00
29		B4	71.40
30		B5	75.00

No.	Local Governments and sub-districts		Probability of flood risk (%)
31	Busan	B6	75.00
32		B7	50.00
33		B8	71.40
34		B9	75.00
35		B10	83.30
36		B11	75.00
37		B12	83.30
38		B13	75.00
39		B14	83.30
40		B15	66.70
41		B16	90.00
42	Daegu	DG1	40.00
43		DG2	90.90
44		DG3	25.00
45		DG4	25.00

No.	Local Governments and sub-districts		Probability of flood risk (%)
46	Daegu	DG5	71.40
47		DG6	33.30
48		DG7	60.00
49		DG8	66.70
50	Incheon	I1	80.00
51		I2	75.00
52		I3	88.90
53		I4	88.90
54		I5	66.70
55		I6	50.00
56		I7	75.00
57		I8	93.80
58		I9	93.80
59		I10	88.90

No.	Local Governments and sub-districts		Probability of flood risk (%)
60	Gwangju	GJ1	33.30
61		GJ2	40.00
62		GJ3	84.60
63		GJ4	40.00
64		GJ5	94.70
65	Daejeon	DJ1	93.80
66		DJ2	88.90
67		DJ3	93.80
68		DJ4	93.80
69		DJ5	88.90
70	Ulsan	U1	83.30
71		U2	83.30
72		U3	94.70
73		U4	92.90
74		U5	92.90

No.	Local Governments and sub-districts		Probability of flood risk (%)
75	Gyeonggi-do	GG1	75.00
76		GG2	93.80
77		GG3	75.00
78		GG4	75.00
79		GG5	75.00
80		GG6	80.00
81		GG7	91.70
82		GG8	80.00
83		GG9	93.80
84		GG10	91.70
85		GG11	75.00
86		GG12	80.00
87		GG13	75.00
88		GG14	93.80
89		GG15	93.80

No.	Local Governments and sub-districts		Probability of flood risk (%)
90	Gyeonggi-do	GG16	75.00
91		GG17	88.90
92		GG18	75.00
93		GG19	93.80
94		GG20	83.30
95		GG21	93.80
96		GG22	94.70
97		GG23	91.70
98		GG24	93.80
99		GG25	90.00
100		GG26	88.90
101		GG27	90.00
102		GG28	91.70
103		GG29	66.70
104		GG30	92.30

No.	Local Governments and sub-districts		Probability of flood risk (%)
105	Gyeonggi-do	GG31	92.30
106	Gangwon-do	GW1	92.90
107		GW2	92.90
108		GW3	92.90
109		GW4	93.80
110		GW5	80.00
111		GW6	88.90
112		GW7	92.30
113		GW8	92.30
114		GW9	80.00
115		GW10	83.30
116		GW11	92.30
117		GW12	92.30
118		GW13	83.30
119		GW14	80.00

No.	Local Governments and sub-districts		Probability of flood risk (%)
120	Gangwon-do	GW15	83.30
121		GW16	92.30
122		GW17	66.70
123		GW18	88.90
124	Chungcheong-do	C1	66.70
125		C2	83.30
126		C3	92.90
127		C4	91.70
128		C5	92.90
129		C6	80.00
130		C7	90.00
131		C8	94.70
132		C9	92.90
133		C10	94.70
134		C11	88.90

No.	Local Governments and sub-districts		Probability of flood risk (%)
135	Chungcheong-do	C12	94.70
136		C13	94.70
137		C14	88.90
138		C15	93.80
139		C16	91.70
140		C17	88.90
141		C18	90.90
142		C19	90.00
143		C20	90.00
144		C21	83.30
145		C22	83.30
146		C23	93.80
147		C24	88.90
148		C25	91.70
149		C26	83.30

No.	Local Governments and sub-districts		Probability of flood risk (%)
150	Chungcheong-do	C27	88.90
151		C28	91.70
152	Jeolla-do	J1	66.70
153		J2	91.70
154		J3	90.90
155		J4	60.00
156		J5	84.60
157		J6	90.90
158		J7	92.90
159		J8	90.00
160		J9	88.90
161		J10	90.00
162		J11	92.90
163		J12	92.90
164		J13	94.70

No.	Local Governments and sub-districts		Probability of flood risk (%)
165	Jeolla-do	J14	91.70
166		J15	71.40
167		J16	94.70
168		J17	66.70
169		J18	90.90
170		J19	84.60
171		J20	84.60
172		J21	92.90
173		J22	92.90
174		J23	94.70
175		J24	84.60
176		J25	92.90
177		J26	92.90
178		J27	94.70
179		J28	94.70

No.	Local Governments and sub-districts		Probability of flood risk (%)
181	Jeolla-do	J30	94.70
182		J31	90.90
183		J32	94.70
184		J33	84.60
185		J34	94.70
186		J35	94.70
187		J36	92.90
188	Gyeongsang-do	GS1	75.00
189		GS2	92.90
190		GS3	80.00
191		GS4	83.30
192		GS5	83.30
193		GS6	88.90
194		GS7	92.90
195		GS8	92.90

No.	Local Governments and sub-districts		Probability of flood risk (%)
196	Gyeongsang-do	GS9	88.90
197		GS10	84.60
198		GS11	83.30
199		GS12	66.70
200		GS13	80.00
201		GS14	92.30
202		GS15	92.30
203		GS16	88.90
204		GS17	80.00
205		GS18	92.90
206		GS19	92.90
207		GS20	84.60
208		GS21	92.30
209		GS22	92.30
210		GS23	80.00

No.	Local Governments and sub-districts		Probability of flood risk (%)
211	Gyeongsang-do	GS24	83.30
212		GS25	84.60
213		GS26	60.00
214		GS27	84.60
215		GS28	66.70
216		GS29	90.90
217		GS30	90.90
218		GS31	83.30
219		GS32	66.70
220		GS33	80.00
221		GS34	80.00
222		GS35	83.30
223		GS36	83.30
224		GS37	84.60
225		GS38	92.90

No.	Local Governments and sub-districts		Probability of flood risk (%)
226	Gyeongsang-do	GS39	92.90
227		GS40	92.90
228		GS41	88.90
229		GS42	88.90
230		GS43	80.00
231	Jeju	JJ1	50.00
232		JJ2	50.00

Appendix C1.

Conditional Probability Tables of Bayesian Network for Seoul

(1) Flood risk (X_1)

Summer precipitation	Population density	Road area ratio	Regional average slope	Days over 80 mm rainfall	low	high
normal	low	normal	low	high	50	50
normal	low	normal	normal	low	50	50
normal	low	normal	normal	high	25	75
normal	low	normal	high	low	50	50
normal	low	normal	high	high	33.333	66.667
normal	low	high	low	low	50	50
normal	low	high	low	high	50	50
normal	low	high	normal	low	50	50
normal	low	high	normal	high	50	50
normal	low	high	high	low	50	50
normal	low	high	high	high	50	50
normal	normal	low	low	low	50	50
normal	normal	low	low	high	66.667	33.333
normal	normal	low	normal	low	33.333	66.667
normal	normal	low	normal	high	50	50
normal	normal	low	high	low	50	50
normal	normal	low	high	high	50	50
normal	normal	normal	low	low	66.667	33.333
normal	normal	normal	low	high	50	50
normal	normal	normal	normal	low	75	25
normal	normal	normal	normal	high	25	75
normal	normal	normal	high	low	33.333	66.667
normal	normal	normal	high	high	50	50
normal	normal	high	low	low	50	50
normal	normal	high	low	high	50	50
normal	normal	high	normal	low	50	50
normal	normal	high	normal	high	33.333	66.667
normal	normal	high	high	low	50	50
normal	normal	high	high	high	40	60
normal	high	low	low	low	50	50
normal	high	low	low	high	66.667	33.333
normal	high	low	normal	low	50	50
normal	high	low	normal	high	50	50
normal	high	low	high	low	50	50
normal	high	low	high	high	50	50
normal	high	normal	low	low	50	50
normal	high	normal	low	high	50	50
normal	high	normal	normal	low	50	50
normal	high	normal	normal	high	33.333	66.667
normal	high	normal	high	low	50	50
normal	high	normal	high	high	50	50

(2) Daily maximum precipitation (X_2)

low	normal	high
28.571	39.286	32.143

(3) 5 days maximum rainfall period (X_3)

Daily maximum precipitation	low	normal	high
low	80	10	10
normal	7.692	84.615	7.692
high	9.091	9.091	81.818

(4) Summer precipitation (June to September) (X_4)

5 days maximum rainfall period	Daily maximum precipitation	Days over 80 mm rainfall	low	normal	high
low	low	low	33.333	33.333	33.333
low	low	high	10	80	10
low	normal	low	33.333	33.333	33.333
low	normal	high	33.333	33.333	33.333
low	high	low	33.333	33.333	33.333
low	high	high	33.333	33.333	33.333
normal	low	low	33.333	33.333	33.333
normal	low	high	33.333	33.333	33.333
normal	normal	low	40	50	10
normal	normal	high	16.667	50	33.333
normal	high	low	33.333	33.333	33.333
normal	high	high	33.333	33.333	33.333
high	low	low	33.333	33.333	33.333
high	low	high	33.333	33.333	33.333
high	normal	low	33.333	33.333	33.333
high	normal	high	33.333	33.333	33.333
high	high	low	25	50	25
high	high	high	10	80	10

(5) Area ratio with the banks (X_5)

Total population	Summer precipitation	Regional average slope	Days over 80 mm rainfall	low	normal	high
low	low	low	low	33.333	33.333	33.333
low	low	low	high	33.333	33.333	33.333
low	low	normal	low	33.333	33.333	33.333
low	low	normal	high	33.333	33.333	33.333
low	low	high	low	33.333	33.333	33.333
low	low	high	high	33.333	33.333	33.333
low	normal	low	low	33.333	33.333	33.333
low	normal	low	high	33.333	33.333	33.333
low	normal	normal	low	50	25	25
low	normal	normal	high	20	60	20
low	normal	high	low	33.333	33.333	33.333
low	normal	high	high	25	50	25
low	high	low	low	33.333	33.333	33.333
low	high	low	high	33.333	33.333	33.333
low	high	normal	low	33.333	33.333	33.333
low	high	normal	high	33.333	33.333	33.333
low	high	high	low	33.333	33.333	33.333
low	high	high	high	33.333	33.333	33.333
normal	low	low	low	33.333	33.333	33.333
normal	low	low	high	33.333	33.333	33.333
normal	low	normal	low	20	60	20
normal	low	normal	high	33.333	33.333	33.333
normal	low	high	low	33.333	33.333	33.333
normal	low	high	high	33.333	33.333	33.333
normal	normal	low	low	25	50	25
normal	normal	low	high	20	40	40
normal	normal	normal	low	20	40	40
normal	normal	normal	high	10	70	20
normal	normal	high	low	50	25	25
normal	normal	high	high	60	20	20
normal	high	low	low	33.333	33.333	33.333
normal	high	low	high	33.333	33.333	33.333
normal	high	normal	low	33.333	33.333	33.333
normal	high	normal	high	33.333	33.333	33.333
normal	high	high	low	33.333	33.333	33.333
normal	high	high	high	33.333	33.333	33.333
high	low	low	low	25	25	50
high	low	low	high	33.333	33.333	33.333
high	low	normal	low	33.333	33.333	33.333
high	low	normal	high	33.333	33.333	33.333
high	low	high	low	33.333	33.333	33.333

(6) Total population (X_6)

Regional average slope	low	normal	high
low	12.5	50	37.5
normal	22.222	66.667	11.111
high	25	50	25

(7) Regional average slope (X_7)

low	normal	high
21.429	57.143	21.429

(8) Road area ratio (X_8)

Population density	Regional average slope	low	normal	high
low	low	33.333	33.333	33.333
low	normal	20	60	20
low	high	25	50	25
normal	low	28.571	42.857	28.571
normal	normal	14.286	71.429	14.286
normal	high	14.286	28.571	57.143
high	low	50	25	25
high	normal	40	40	20
high	high	33.333	33.333	33.333

(9) Population density (X_9)

Total population	low	normal	high
low	42.857	42.857	14.286
normal	10	70	20
high	14.286	71.429	14.286

(10) Capacity of drainage facilities (X_{10})

Summer precipitation	Population density	Road area ratio	Regional average slope	Days over 80 mm rainfall	low	high
low	low	low	low	low	50	50
low	low	low	low	high	50	50
low	low	low	normal	low	50	50
low	low	low	normal	high	50	50
low	low	low	high	low	50	50
low	low	low	high	high	50	50
low	low	normal	low	low	50	50
low	low	normal	low	high	50	50
low	low	normal	normal	low	50	50
low	low	normal	normal	high	50	50
low	low	normal	high	low	50	50
low	low	normal	high	high	50	50
low	low	high	low	low	50	50
low	low	high	low	high	50	50
low	low	high	normal	low	50	50
low	low	high	normal	high	50	50
low	low	high	high	low	50	50
low	low	high	high	high	50	50
low	normal	low	low	low	50	50
low	normal	low	low	high	50	50
low	normal	low	normal	low	50	50
low	normal	low	normal	high	50	50
low	normal	low	high	low	50	50
low	normal	low	high	high	50	50
low	normal	normal	low	low	50	50
low	normal	normal	low	high	50	50
low	normal	normal	normal	low	66.667	33.333
low	normal	normal	normal	high	50	50
low	normal	normal	high	low	50	50
low	normal	normal	high	high	50	50
low	normal	high	low	low	66.667	33.333
low	normal	high	low	high	50	50
low	normal	high	normal	low	50	50
low	normal	high	normal	high	50	50
low	normal	high	high	low	50	50
low	normal	high	high	high	50	50
low	high	low	low	low	50	50
low	high	low	low	high	50	50
low	high	low	normal	low	66.667	33.333
low	high	low	normal	high	50	50
low	high	low	high	low	50	50

(11) Days over 80 mm rainfall (X_{11})

Daily maximum precipitation	low	high
low	11.111	88.889
normal	66.667	33.333
high	20	80

Appendix C2.

Conditional Probability Tables of Bayesian Network for Korea

(1) Flood risk (X_1)

Summer precipitation	Number of civil servants (water)	Capacity of drainage facilities	Area ratio with the banks	Road area ratio	low	high
low	low	low	low	low	50	50
low	low	low	low	normal	20	80
low	low	low	low	high	25	75
low	low	low	normal	low	33.333	66.667
low	low	low	normal	normal	11.111	88.889
low	low	low	normal	high	6.25	93.75
low	low	low	high	low	50	50
low	low	low	high	normal	50	50
low	low	low	high	high	33.333	66.667
low	low	high	low	low	33.333	66.667
low	low	high	low	normal	50	50
low	low	high	low	high	50	50
low	low	high	normal	low	50	50
low	low	high	normal	normal	16.667	83.333
low	low	high	normal	high	8.333	91.667
low	low	high	high	low	50	50
low	low	high	high	normal	50	50
low	low	high	high	high	50	50
low	high	low	low	low	50	50
low	high	low	low	normal	50	50
low	high	low	low	high	25	75
low	high	low	normal	low	50	50
low	high	low	normal	normal	50	50
low	high	low	normal	high	11.111	88.889
low	high	low	high	low	50	50
low	high	low	high	normal	50	50
low	high	low	high	high	33.333	66.667
low	high	high	low	low	50	50
low	high	high	low	normal	50	50
low	high	high	low	high	50	50
low	high	high	normal	low	50	50
low	high	high	normal	normal	50	50
low	high	high	normal	high	33.333	66.667
low	high	high	high	low	50	50
low	high	high	high	normal	50	50
low	high	high	high	high	20	80
normal	low	low	low	low	7.692	92.308
normal	low	low	low	normal	10	90
normal	low	low	low	high	33.333	66.667
normal	low	low	normal	low	16.667	83.333
normal	low	low	normal	normal	7.143	92.857

(2) Daily maximum precipitation (X_2)

Daily maximum precipitation	low	normal	high
low	72.222	25	2.778
normal	9.174	88.991	1.835
high	1.449	49.275	49.275

(3) 5 days maximum rainfall period (X_3)

Daily maximum precipitation	low	normal	high
low	72.222	25	2.778
normal	9.174	88.991	1.835
high	1.449	49.275	49.275

(4) Summer precipitation (June to September) (X_4)

Days over 80 mm rainfall	Lowland area of less than el. 10	Daily maximum precipitation	5 days maximum rainfall period	low	normal	high
low	low	low	low	33.333	33.333	33.333
low	low	low	normal	33.333	33.333	33.333
low	low	low	high	33.333	33.333	33.333
low	low	normal	low	33.333	33.333	33.333
low	low	normal	normal	22.222	66.667	11.111
low	low	normal	high	33.333	33.333	33.333
low	low	high	low	33.333	33.333	33.333
low	low	high	normal	6.25	75	18.75
low	low	high	high	7.143	21.429	71.429
low	high	low	low	33.333	33.333	33.333
low	high	low	normal	33.333	33.333	33.333
low	high	low	high	33.333	33.333	33.333
low	high	normal	low	60	20	20
low	high	normal	normal	13.636	31.818	54.545
low	high	normal	high	25	25	50
low	high	high	low	33.333	33.333	33.333
low	high	high	normal	10	50	40
low	high	high	high	4.167	66.667	29.167
high	low	low	low	83.333	8.333	8.333
high	low	low	normal	50	25	25
high	low	low	high	33.333	33.333	33.333
high	low	normal	low	60	20	20
high	low	normal	normal	60	30	10
high	low	normal	high	33.333	33.333	33.333
high	low	high	low	33.333	33.333	33.333
high	low	high	normal	60	20	20
high	low	high	high	33.333	33.333	33.333
high	high	low	low	73.684	21.053	5.263
high	high	low	normal	30	60	10
high	high	low	high	33.333	33.333	33.333
high	high	normal	low	62.5	25	12.5
high	high	normal	normal	26.866	43.284	29.851
high	high	normal	high	33.333	33.333	33.333
high	high	high	low	33.333	33.333	33.333
high	high	high	normal	25	25	50
high	high	high	high	25	50	25

(5) Area ratio with the banks (X_5)

Lowland area of less than el. 10	Total population	low	normal	high
low	low	13.636	77.273	9.091
low	normal	9.524	80.952	9.524
low	high	17.647	70.588	11.765
high	low	42.593	48.148	9.259
high	normal	32.143	60.714	7.143
high	high	24.528	64.151	11.321

(6) Total population (X_6)

Regional average slope	low	normal	high
low	15	30	55
normal	28.788	40.152	31.061
high	69.048	21.429	9.524

(7) Regional average slope (X_7)

Lowland area of less than el. 10	low	normal	high
low	35.185	55.556	9.259
high	12.739	64.331	22.93

(8) Road area ratio (X_8)

Regional average slope	Population density	GRDP	low	normal	high
low	low	low	33.333	33.333	33.333
low	low	normal	25	25	50
low	low	high	12.5	37.5	50
low	normal	low	25	25	50
low	normal	normal	16.667	16.667	66.667
low	normal	high	12.5	12.5	75
low	high	low	20	20	60
low	high	normal	7.692	7.692	84.615
low	high	high	7.692	7.692	84.615
normal	low	low	33.333	33.333	33.333
normal	low	normal	10.526	42.105	47.368
normal	low	high	9.615	50	40.385
normal	normal	low	20	20	60
normal	normal	normal	14.286	28.571	57.143
normal	normal	high	3.226	38.71	58.065
normal	high	low	14.286	14.286	71.429
normal	high	normal	12.5	12.5	75
normal	high	high	4.167	4.167	91.667
high	low	low	33.333	33.333	33.333
high	low	normal	50	45.833	4.167
high	low	high	41.176	52.941	5.882
high	normal	low	33.333	33.333	33.333
high	normal	normal	20	40	40
high	normal	high	25	50	25
high	high	low	25	25	50
high	high	normal	33.333	33.333	33.333
high	high	high	33.333	33.333	33.333

(9) Population density (X_9)

Total population	low	normal	high
low	90.411	6.849	2.74
normal	52.703	27.027	20.27
high	5.97	35.821	58.209

(10) Capacity of drainage facilities (X_{10})

GRDP	Regional average slope	Road area ratio	Daily maximum precipitation	low	high
low	low	low	low	50	50
low	low	low	normal	50	50
low	low	low	high	50	50
low	low	normal	low	50	50
low	low	normal	normal	50	50
low	low	normal	high	50	50
low	low	high	low	50	50
low	low	high	normal	80	20
low	low	high	high	50	50
low	normal	low	low	50	50
low	normal	low	normal	50	50
low	normal	low	high	50	50
low	normal	normal	low	50	50
low	normal	normal	normal	50	50
low	normal	normal	high	50	50
low	normal	high	low	66.667	33.333
low	normal	high	normal	85.714	14.286
low	normal	high	high	50	50
low	high	low	low	50	50
low	high	low	normal	50	50
low	high	low	high	50	50
low	high	normal	low	50	50
low	high	normal	normal	50	50
low	high	normal	high	50	50
low	high	high	low	50	50
low	high	high	normal	66.667	33.333
low	high	high	high	50	50
normal	low	low	low	50	50
normal	low	low	normal	50	50
normal	low	low	high	50	50
normal	low	normal	low	50	50
normal	low	normal	normal	50	50
normal	low	normal	high	50	50
normal	low	high	low	80	20
normal	low	high	normal	50	50
normal	low	high	high	33.333	66.667
normal	normal	low	low	66.667	33.333
normal	normal	low	normal	50	50
normal	normal	low	high	50	50
normal	normal	normal	low	75	25
normal	normal	normal	normal	75	25

(11) Days over 80 mm rainfall (X_{11})

Daily maximum precipitation	low	high
low	2.857	97.143
normal	26.852	73.148
high	92.647	7.353

(12) Lowland area of less than el. 10 (X_{12})

low	high
25.121	74.879

(13) Number of civil servants related to water (X_{13})

Summer precipitation	GRDP	Total population	Area ratio with the banks	Lowland area of less than el. 10	low	high
low	low	low	low	low	50	50
low	low	low	low	high	50	50
low	low	low	normal	low	50	50
low	low	low	normal	high	50	50
low	low	low	high	low	50	50
low	low	low	high	high	50	50
low	low	normal	low	low	50	50
low	low	normal	low	high	50	50
low	low	normal	normal	low	50	50
low	low	normal	normal	high	50	50
low	low	normal	high	low	50	50
low	low	normal	high	high	50	50
low	low	high	low	low	50	50
low	low	high	low	high	50	50
low	low	high	normal	low	50	50
low	low	high	normal	high	57.143	42.857
low	low	high	high	low	50	50
low	low	high	high	high	50	50
low	normal	low	low	low	50	50
low	normal	low	low	high	66.667	33.333
low	normal	low	normal	low	80	20
low	normal	low	normal	high	75	25
low	normal	low	high	low	50	50
low	normal	low	high	high	50	50
low	normal	normal	low	low	50	50
low	normal	normal	low	high	33.333	66.667
low	normal	normal	normal	low	50	50
low	normal	normal	normal	high	80	20
low	normal	normal	high	low	33.333	66.667
low	normal	normal	high	high	50	50
low	normal	high	low	low	50	50
low	normal	high	low	high	33.333	66.667
low	normal	high	normal	low	60	40
low	normal	high	normal	high	25	75
low	normal	high	high	low	33.333	66.667
low	normal	high	high	high	66.667	33.333
low	high	low	low	low	50	50
low	high	low	low	high	75	25
low	high	low	normal	low	75	25
low	high	low	normal	high	50	50
low	high	low	high	low	50	50

(14) GRDP (X_{14})

Total population	Population density	low	normal	high
low	low	1.471	39.706	58.824
low	normal	14.286	28.571	57.143
low	high	25	25	50
normal	low	2.439	26.829	70.732
normal	normal	4.545	22.727	72.727
normal	high	17.647	23.529	58.824
high	low	16.667	50	33.333
high	normal	15.385	19.231	65.385
high	high	14.634	31.707	53.659

초 록

기후변화로 인한 기상이변은 국내뿐만 아니라 전 세계적으로도 가장 큰 관심사 중에 하나로 한파와 폭설, 집중호우와 홍수, 가뭄과 산불과 같은 자연재해가 전 세계적으로 발생하였다. 실제로 국내의 경우 기상이변으로 인하여 한반도 내 국지적 집중호우가 빈발하고 서울 도심이 침수되는 등 많은 지역들이 홍수에 노출되는 위험도가 증가되고 있다. 이에 따라 치수 관련 정보에 대한 국민들의 관심이 크게 증가하고 있으며 물 관리에 대한 대중의 참여증대와 환경에 대한 인식의 제고 등으로 치수관련 정책의 정당성에 대한 설득력이 뒷받침되어야 하는 상황이다. 따라서 기상이변으로 인한 자연재해에 대응하기 위한 대책 수립이 중요하며 치수 관련 계획 및 사업의 우선순위 선정에 활용할 수 있는 다양한 분석 정보도 제공할 수 있는 체계를 구축해야 한다.

본 논문에서는 우리 나라의 대표적인 자연재해에 속하는 홍수에 대한 위험도 분석을 실시하였다. 홍수에 대한 우리 사회의 적응 시스템의 특성이 반영될 수 있도록 기후특성뿐 아니라 사회•경제적인 특성 간의 복잡한 인과관계를 고려하기 위하여 베이지안 네트워크(Bayesian networks)를 적용하였다. 베이지안

네트워크는 여러 의존관계와 불확실성을 포함한 복잡한 문제를 표현하고 분석할 수 있는 방법으로, 홍수 위험도와 관계된 대용변수들의 의존관계를 모형화하고 분석할 수 있는 틀을 제공한다. 또한 베이지안 네트워크는 전문가의 의견이나 문헌 등과 같은 선험적 지식을 충분히 활용할 수 있어 객관적인 정보를 획득하기 어려운 홍수위험도를 분석하기에 효과적이다.

현재 홍수로 인한 피해는 인명이나 재산피해로 집계되고 있기 때문에 홍수피해금액을 추산하여 홍수위험도를 평가하였고 이를 위해 12개 광역시도를 대상으로 전국에 대하여 네트워크를 구성하였다. 특별히 서울의 경우 타 지역과 차별화된 인문, 사회적 특성을 고려하여 별도로 네트워크를 구성하였다. 또한 홍수위험도 분석을 위해 선정된 대용변수 입력자료의 불확실성을 해소하기 위해 원자료(raw data)를 사용하였다.

본 논문에서 제안한 방법은 홍수위험도 모형화, 홍수 위험도 분석, 홍수 위험도 진단의 세 단계로 구성된다. 첫 번째로, 홍수 위험도 모형화 단계에서는 베이지안 네트워크를 구성하여 홍수 위험도에 영향을 미치는 변수들의 의존관계를 정성적으로 나타낸 후, 의존관계의 정도를 조건부 확률을 이용하여 정량적으로 나타낸다. 두 번째 단계에서는 구축된 베이지안 네트워크를

활용하여 홍수 위험도를 분석할 수 있는 두 가지 방안이 제안된다. 먼저, 지역별 홍수 위험도 확률을 도출하고 베이지안 네트워크 결과의 우수성과 타당성을 입증하기 위하여 실제홍수피해와 비교분석한다. 한편으로, 홍수 위험도 분석을 통해 특정 변수의 영향력을 고찰하고 이를 정량화할 수 있는 결과를 제시한다. 마지막으로 홍수 위험도 진단을 위해 도출된 결과를 토대로 전국에 대한 홍수위험도 지도(flood risk mapping)를 제시하여 치수관련 계획에 유용한 정보를 제공할 수 있는 토대를 마련한다.

모형화 결과 서울에서는 도로면적비율, 총인구, 여름강수량 대용변수가 홍수에 대해 큰 민감도를 보였다. 전국에서는 물관리공무원수, 제방사용면적율, 여름강수량, 내수배제시설 배수능력 대용변수가 큰 민감도를 나타냈다. 높은 민감도를 보이는 위의 대용변수 값이 클수록 홍수위험도가 높게 나타났으며 대전, 충청도, 울산 등이 높은 홍수위험도를 보였다. 또한 부산, 대구, 광주, 대전과 같은 대도시의 홍수위험도 결과가 실제홍수피해와 일치되는 결과를 보여준다.

본 연구에서 제안한 방법은 홍수위험도의 모형화, 분석, 진단을 포괄적으로 다룸으로써 홍수 위험도 관리를 효과적으로 지원할 수 있을 것으로 기대된다. 특히 불확실성을 고려한 지역별

홍수 위험도 분석 결과는 최적의 홍수방어 대책 수립을 위한 기초자료로 유용할 것으로 기대된다.

주요어: 베이지안 네트워크, 홍수위험 분석, 홍수위험지도

학 번: 2010-23308